# Convergence Analysis for Finite Element Discretizations of Highly Indefinite Problems

**Dissertation**

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr.sc.nat)

vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von

**Asieh Parsania**

aus dem Iran

**Promotionskomitee**

Prof. Dr. Stefan A. Sauter (Vorsitz)
Prof. Dr. Michel Chipot

Zürich, 2012

# Abstract

Helmholtz problem appears in many areas for example in the context of inverse and scattering problems. This problem is solved numerically and the challenge is that the solutions become highly oscillatory. As a consequence the numerical discretization has to be adapted to resolve these oscillations. Galerkin methods are well established to solve elliptic problems - however for Helmholtz problems they suffer from the indefiniteness of the equation, more precisely, the stability of the discrete solution as well as the corresponding error is significantly "polluted" in the preasymptotic range.

For high wave numbers, the solution shows a non-robust behavior which is known as the pollution effect, i.e. the discrepancy between the best approximation error and the error of the Galerkin solution increases with increasing wave number. The physical reason for this is the highly oscillatory nature of the solution of this problem while in the mathematical language the Helmholtz equation becomes highly indefinite with increasing wave number. It is an important topic of research in numerical analysis to find an efficient numerical discretization which behaves reasonably robust with respect to the wave number.

One of the interesting questions in this area is how the performance of the method can be affected by the parameters like the wave number and the mesh size. Classical conforming low-order finite elements suffer from pollution effect. The minimal dimension, $N$, e.g., of the $\mathcal{P}_1$-finite elements must satisfy $N \gtrsim k^{2d}$ where $d$ is spatial dimension. Previous works show that the high oscillation of the solution can be resolved by refining the mesh size, $h$, and the pollution effect can be reduced by employing higher order methods with comparison to lower order methods. For example it is known that for conforming finite elements we can get a more relaxed condition ($N \gtrsim k^d$) if we use higher order methods. In the recent years much progress has been made for non-conforming finite element discretizations of the Helmholtz problem. Among them are the plane wave discretization in combination with the ultra-weak variational formulation which turn out to be unconditionally stable.

In this thesis, we develop a stability and convergence theory for the Ultra Weak Variational Formulation (UWVF) of a highly indefinite Helmholtz problem in $\mathbb{R}^d, d = 1,2,3$ for general, abstract trial and test spaces. The theory covers conforming as

well as nonconforming generalized finite element methods. In contrast to conventional Galerkin methods where a minimal resolution condition is necessary to guarantee the unique solvability, it is proved that the UWVF admits a unique solution without any condition for piecewise polynomials and plane waves and under a mild condition for a very general class of the approximation spaces.

We develop a theory for general abstract non-conforming Galerkin discretization. As an application we present the error analysis for the conforming and non-conforming $hp$-version of the finite element method explicitly in terms of the mesh width $h$, polynomial degree $p$ and wave number $k$ for two different cases. We show that our method converges with almost optimal order under the conditions that $kh/p$ is sufficiently small and the polynomial degree $p$ is at least $O(\log k)$.

# Zusammenfassung

Das Helmholtz-Problem kommt in vielen Gebieten vor, beispielsweise im Zusammenhang mit Streuungsproblemen. Unser ziel ist es, dieses Problem numerisch zu lösen. Die Hauptschwierigkeit ist hierbei, dass die Lösungen stark oszillieren. Infolgedessen muss die numerische Diskretisierung angepasst werden, um diese Oszillationen aufzulösen. Galerkin-Methoden haben sich für die Lösung elliptischer Probleme gut etabliert. Für Helmholtz-Probleme entstehen jedochstabilitätsprobleme aufgrund der Indefinitheit der Gleichung, genauer, die Stabilität der diskreten Lösung sowie der zugehörige Fehler werden im präasymptotischen Bereich signifikant“gestört”.

Für grosse Wellenzahlen zeigt die Lösung ein nichtrobustes Verhalten, welches als Pollution-Effekt bekannt ist, d.h. die Diskrepanz zwischen dem Fehler der bestmöglichen Approximation und dem Fehler der Galerkin-Lösung wächst mit wachsender Wellenzahl. Der physikalische Grund dafür ist die stark oszillierende Natur der Lösung dieses Problems. Der “mathematische” Grund besteht darin, dass die Helmholtz-Gleichung mit wachsender Wellenzahl zunehmend indefinit wird. Es ist ein wichtiges Forschungsgebiet der numerischen Analysis, eine effiziente numerische Diskretisierung zu finden, welche sich robust verhält bezüglich der Wellenzahl.

Eine der interessanten Fragen auf diesem Gebiet ist, wie die Methode durch Parameter wie die Wellenzahl und die Maschenweite beeinflusst werden kann. Klassische konforme Finite Elemente niedriger Ordnung sind wegen des Pollution-Effekts ungeeignet. Die minimale Dimension $N$, z.B. der $\mathcal{P}_1$-Finite Elemente, muss $N \gtrsim k^{2d}$ erfüllen, wobei $d$ die Raumdimension ist. Frühere Arbeiten zeigen, dass die starke Oszillation der Lösung mittels Verfeinerung der Maschenweite $h$ aufgelöst werden kann, und der Pollution-Effekt kann durch Anwendung von Methoden höherer Ordnung reduziert werden im Vergleich zu Methoden niedrigerer Ordnung. Es ist beispielsweise bekannt, dass wir für konforme Finite Elemente eine schwächere Bedingung ($N \gtrsim k^d$) erhalten können, wenn wir Elemente höherer Ordnung benutzen. In den letzten Jahren wurden bei den nichtkonformen Finite Elemente Diskretisierungen des Helmholtz Problems viele Fortschritte gemacht. Darunter sind die Plane-Wave-Diskretisierungen in Kombination mit der “Ultra Weak Variational Formulation”, welche sich als absolut stabil erweisen.

In dieser Dissertation entwickeln wir eine Stabilitäts- und Konvergenztheorie für die ultraschwache Variationsformulierung (UWVF) eines stark indefiniten Helmholtz Problems in $\mathbb{R}^d, d = 1,2,3$. Die Theorie deckt sowohl konforme als auch nichtkonforme verallgemeinerte Finite-Elemente-Methoden ab. Im Gegensatz zu konventionellen Galerkin-Methoden, bei welchen eine minimale Auflösungsbedingung notwendig ist, um die eindeutige Lösbarkeit zu garantieren, wird bewiesen, dass die UWVF für stückweise Polynome und für ebene Wellen ohne irgendeine Bedingung sowie unter einer schwachen Bedingung für eine sehr allgemeine Klasse von Approximationsrämen eine eindeutige Lösung liefert.

Wir entwickeln eine Theorie für allgemeine abstrakte nichtkonforme Galerkin-Diskretisierungen. Als Anwendung präsentieren wir die Fehleranalyse für die *hp*-Version der Finite Elemente Methode explizit für die Maschenweite $h$, den Polynomgrad $p$ und die Wellenzahl $k$ für zwei verschiedene Fälle. Wir zeigen, dass unsere Methode unter den Bedingungen, dass $kh/p$ genügend klein ist und dass der Polynomgrad $p$ mindestens $O(\log k)$ ist, optimale Konvergenzordnung besitzt.

# Acknowledgements

First and foremost I want to thank my Ph.D. supervisor, Professor Dr. Stefan Sauter to be supportive since the days I began to work on this project. I sincerely appreciate all his contributions of time, ideas and funding in this period.

I am also very grateful to Professor Dr. Michel Chipot for all his help and encouragement.

I would like to thank Professor Dr. Markus Melenk for his scientific collaborations and support during my short visit of TU Wien.

Special mention must be made of those professors who serve as chairs and members of dissertation committees.

I also would like to thank all my friends in Zürich who made these 3.5 years unforgettable for me. All of my friends at institute of Mathematics; my early friends whom we started to learn German and our first trips in Switzerland together: Viorica and Utsav; My Iranian friends, with whom I didn't feel myself far from home.

I am greatly indebted to my best friend Jafar, for his endless support during the time of the writing the thesis.

Last, but by no means least, I thank my whole family for all their support, patience and encouragement over all these years.

Finally, I am glad to acknowledge the different source of financial support during this period; namely, Universität Zürich and Swiss National science foundation (SNSF) Nr. 124825.

*Dedicated to my parents,*
*thank you for all of your love, support and guidance throughout my life.*

# Table of Contents

# List of Symbols

# 1

# Introduction

Waves, as physical phenomena, can be defined as “a disturbance or variation that transfers energy progressively from point to point in a medium and that may take the form of an elastic deformation or of a variation of pressure, electric or magnetic intensity, electric potential, or temperature.”

For the mathematically formulation of wave motion one employs the concept of a wave function, which describes the position of a particle in the medium at any time. The most basic prototype of a wave function is the sine wave or sinusoidal wave, which is a periodic wave. It is important to note that the wave function represents the displacement about the equilibrium position. Some of the properties of the wave function are

- **Wave speed** - the speed of the wave’s propagation,
- **Amplitude** - the maximum magnitude of the displacement from equilibrium,
- **Period** - the time for one wave cycle,
- **Frequency** - the number of cycles in a unit of time,
- **Wave length** - the distance between any two points at corresponding positions on successive repetitions in the wave,
- **Wave number** ($k$) - $2\pi$ divided by the wavelength.

Waves are everywhere. Most of the information that we receive every day comes in the form of waves, e.g. radio, TV, music. A Tsunamis or tidal wave is a large water wave that is produced by some kind of seismic phenomena. Shock waves created by a lighting may be a sonic boom. Sonic booms can be produced by aircraft flying at speeds greater than the speed of sound in air. The massive compression waves produced by an earthquake are similar to sound waves. The quality of sound coming from a musical instrument depends upon the number of harmonic frequencies produced and their relative intensities. Also in radar communications and Ultrasonic applications and electromagnetic are also some other examples of wave applications. These examples show that why it is important to study waves.

From the numerical point of view, the wave propagation problems with high frequencies (or large wave numbers) are very challenging. The highly oscillatory nature of the solutions of such kind of problems result in a strong singular perturbation of the elliptic problem. The difficulty is to establish a robust numerical method with a reasonable mesh constraint in the finite element method (FEM) and to understand the dependence of the accuracy of the method on the discretization and problem parameters which include the geometry of the domain and in particular the mesh width. Therefore, it is essential that the numerical model can resolve the difficulties associated with large wave number.

Highly indefinite boundary value problems arise for example, in conditions that electromagnetic or acoustic scattering problems are modeled in frequency domain. There the Helmholtz equation with a high wave number $k$ is an adequate model problem. The Helmholtz equation belongs to the classical equations of mathematical physics. The existence and uniqueness of solutions of this equation was studied in 1950's [12, 41].

The development of the finite element method (FEM) for acoustic problems goes back to the 1970's [13, 33, 34, 53, 60]. This method has been for many years used to discretize the different types of Helmholtz problems. By considering the polynomial approximation of order $p$ of an oscillatory wave, e.g., $\sin kx$ on a small interval of length $h$ one easily derives the condition $kh/p < 1$ for a minimal resolution of the wave. However, it was proved that the quasi optimality of the finite element error estimates can be obtained under the stronger mesh constraint $k^2 h \lesssim 1$ [4, 19, 20]. It is also known that the unique solvability of the low order $h$-version finite element methods is only obtained under a very restrictive stability condition. For example as in [5] for $\mathcal{P}_1$-element space, the condition $N \gtrsim k^{2d}$ must be satisfied, where $N$ is the number of degrees of freedom and $d$ is the spatial dimension $d \in \{1,2,3\}$.

Higher order finite elements, where the polynomial degree is increased logarithmically with respect to the wave number, perform much better than low order finite elements in the pre-asymptotic regime [46, 47]. However, a minimal resolution condition for the finite element space has to be satisfied in order to guarantee the existence of a discrete solution. For example for the case of the domains with analytic boundary the following condition is sufficient to ensure the quasi optimality of the Galerkin method for $hp$-FEM:

$$\frac{kh}{p} \lesssim 1 \qquad \text{together with} \qquad p \gtrsim \log k,$$

where $p$ denotes the order of the method. An important tool in the theory is a $k$-explicit regularity theory from [46, 47] that is based on decomposing the solution $u$ into two parts $u_{H^2} \in H^2$ with $k$-independent regularity constants and the analytic part $u_{\mathcal{A}}$ with $k$-explicit bounds for all derivatives for a right-hand side $f \in L^2(\Omega)$.

In recent decades, in order to minimize or eliminate the pollution effect and ob-

tain a more stable scheme for large wave numbers, new types of methods were developed to solve this problem numerically. There are mainly two groups of methods, the first group is based on variational formulations other than the classical Galerkin methods. Some examples of this group are Galerkin Least Square Finite Element Methods [11, 30, 31], Quasi stabilized Finite Element Methods [7] and Discontinuous Galerkin Methods (DG methods) [2, 8, 9, 10, 23, 24, 28, 32, 42, 48]. The second group is based on numerical methods on non-standard ansatz functions. From this group we recall the stabilized or nonstandard methods such as Generalized Galerkin/Finite Element Methods and Partition of Unity Methods [6, 39, 38, 44, 43, 51, 52, 57]. In addition many discrete approximation spaces have been proposed. Between these methods DG methods are known as very powerful tool for solving partial differential equations.

The main advantages of the DG methods compared to classical conforming Galerkin methods are as follows:

- Since the (weak) continuity is enforced by the DG variational formulation there is a lot of freedom to choose trial and test spaces (even in an element-by-element fashion).
- DG methods can easily handle meshes with hanging nodes and elements of general shapes.
- DG methods have the advantage that the elements can be subdivided independently and hanging nodes do not pose a problem ($h$-refinement). The same is true with adjusting the polynomial order ($p$-refinement).

The Ultra Weak Variational Formulation (UWVF) of Cessenat and Després [9, 10, 16] belongs to the category of the DG methods and became a very popular method in recent years. It allows local discretization spaces as (e.g., plane waves [28, 35]) which are discontinuous and continuity is enforced by the discrete equations in a weak way.

In this dissertation we use the ultra weak variational formulation which was developed in [9, 16, 28] for the Helmholtz problem. Our goal is to develop a theory for ultra weak variational formulation which allows us to derive the quasi-optimal convergence behavior of abstract conforming and non-conforming generalized finite element spaces from certain local approximation properties and local inverse estimates.

The thesis is structured as follows. In Chapter 2, we recall some definitions and properties of the acoustic waves, e.g., conservation laws, and the basic theory of the acoustic waves. In the next step, we discuss some typical boundary conditions for the Helmholtz problem and their physical meaning. We continue this chapter by presenting the weak form of our model problem and the standard finite element formulation of it, followed by the stability and convergence results from [46, 47]. We finish this chapter by recalling the decomposition lemma for the solution $u$ of the Helmholtz problem.

Chapter 3 presents the construction of the ultra weak variational formulation for the Helmholtz problem.

Chapter 4 is at the heart of this thesis. It begins with a discussion of the continuity and

coercivity of the DG sesquilinear form and is followed by an analysis of the existence and uniqueness of the solution of the DG problem. At this point we introduce the essential condition for a discrete space to admit the stability. In the next step we discuss, in detail, the convergence estimates for very general non-conforming finite element spaces. We will derive the optimal-order estimate in the DG norm.

Chapter 5 provides a brief overview of the discrete spaces, e.g., plane waves and piecewise polynomials. We also present some trace estimates.

Chapter 6 contains some applications of our theory. We use our stability and convergence estimates for the cases of general conforming and non-conforming *hp*-finite elements. We also study the approximation property for each of those cases and estimate it explicitly in terms of the wave number $k$, mesh size $h$ and the polynomial degree $p$.

In the last chapter, we compare the results for plane wave discontinuous Galerkin method with our method.

# 2

# Helmholtz Problem

In this chapter we will present an introduction to acoustic waves and in particular to the Helmholtz equation and its different boundary conditions along their physical applications. We introduce the variational problem and the existence and uniqueness of the solution of the Helmholtz problem with Robin boundary condition. In the last section we present the standard Galerkin finite element discretization and discuss about the stability and convergence analysis.

## 2.1 Acoustic Waves

In this section we will give a short introduction to the theory of the acoustic wave propagation. Here we restrict ourselves to the case of fluid or gas medium (if the medium is solid one has to apply the theory of elastic waves). For a more detailed introduction to waves we refer to [1, 3, 25, 40, 49, 54]. We start with an introduction to the fundamental laws: (cf. [36])

(i) Conservation of Mass
Consider the medium (fluid) with pressure $P(\mathbf{x},t)$, density $\rho(\mathbf{x},t)$ and particle velocity $\mathbf{V}(\mathbf{x},t)$, where $t$ denotes the time. The conservation of mass can be stated as follows

$$-\frac{\partial}{\partial t}\int_V \rho dV = \int_{\partial V} \rho(\mathbf{V}\cdot\mathbf{n})dS,$$

where $V$ denotes the volume element and $\partial V$ denotes its boundary and $\mathbf{n}$ is the outward normal vector to the $V$ as it is shown in the Figure 2.1.

Then, from the Gauss theorem we derive the continuity equation

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho\mathbf{V}) = 0. \tag{2.1}$$

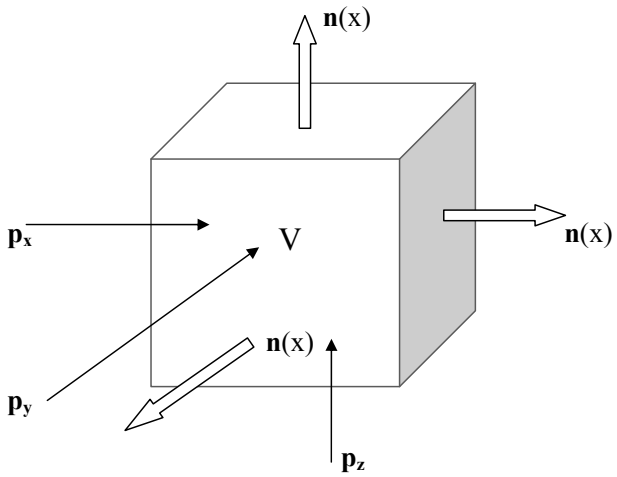


**Figure 2.1**: A volume element with Pressure and normal vectors

(ii) Equation of Motion

Equation of motion or Euler equation states

$$\rho \frac{\partial \mathbf{V}}{\partial t} = -\nabla P. \tag{2.2}$$

From linear material law, we have

$$P = c^2 \rho \tag{2.3}$$

where $c$ is the speed of sound.

Now we assume that the medium is homogeneous, from (2.3) we get

$$\frac{\partial^2 P}{\partial t^2} = c^2 \frac{\partial^2 \rho}{\partial t^2},$$

we combine this result with (2.1) to obtain

$$\begin{aligned} -\frac{1}{c^2}\frac{\partial^2 P}{\partial t^2} &= \frac{\partial}{\partial t}\mathrm{div}(\rho \mathbf{V}) \\ &= \mathrm{div}(\frac{\partial \rho}{\partial t}\mathbf{V} + \rho \frac{\partial \mathbf{V}}{\partial t}) \\ &= \mathrm{div}(\frac{\partial \rho}{\partial t}\mathbf{V}) + \mathrm{div}(\rho \frac{\partial \mathbf{V}}{\partial t}), \end{aligned}$$

from equation of motion it follows

$$\begin{aligned} -\frac{1}{c^2}\frac{\partial^2 P}{\partial t^2} &= \mathrm{div}(\frac{\partial \rho}{\partial t}\mathbf{V}) - \mathrm{div}(\nabla P) \\ &= \mathrm{div}(\frac{\partial \rho}{\partial t}\mathbf{V}) - \Delta P. \end{aligned}$$

Now if we assume that $\rho$ is a constant then we derive the wave equation

$$\Delta P - \frac{1}{c^2}\frac{\partial^2 P}{\partial t^2} = 0.$$

Assuming time harmonic waves, i.e.,

$$F(\mathbf{x},t) = f(\mathbf{x})e^{-i\omega t}$$

with circular frequency $\omega$ and $i = \sqrt{-1}$ denoting the imaginary unit, we obtain the Helmholtz equation

$$\Delta P + k^2 P = 0$$

where the wave number $k$ is defined by $k := \omega/c$.

## 2.2 Helmholtz Equation and Boundary Conditions

The Helmholtz equation is an elliptic equation. In order to obtain a well posed problem we have to formulate suitable boundary conditions. These conditions comes from some physical laws which are formulated on the boundaries of the domain of the problem. Typically the Helmholtz problem is considered in an unbounded exterior domain with hard or soft scatterers at the boundary and the Sommerfeld radiation condition at infinity. For numerical purposes, it is more convenient to formulate this problem on a bounded domain (with artificial boundary) instead of the unbounded domain. This can be done via the Dirichlet-to-Neumann operator (see [50]).

### 2.2.1 Boundary Conditions

(i) **Sommerfeld Radiation Condition**

To guarantee a unique solution for wave problems on unbounded domains, it is necessary to have a condition which represent the behavior of the wave at infinity. As is typical for wave propagation in unbounded acoustic domains (free space) we assume that no waves are reflected from infinity.

Let $u(\mathbf{r})$ be the solution of homogeneous Helmholtz equation in an exterior domain $\Omega^+ = \mathbb{R}^d \setminus \overline{\Omega}$. Assume that a wave source is placed at the origin and let $R$ be the radial distance from the origin to the observation point.

In order to absorb the waves at infinity we impose the following condition (the so called Sommerfeld condition)

$$\lim_{R\to\infty} R^{(d-1)/2}\left(\frac{\partial u}{\partial R} - iku\right) = 0.$$

(ii) **Dirichlet Boundary Condition**

This condition applies in bounded domains (e.g., $\Omega$) when the material of the surface has much lower impedance[1] than the carrier medium, i.e. $\rho_2 v_2 \ll \rho_1 v_1$ where $\rho_1$ and $v_1$ denote respectively the surface medium density and the propagation velocity of ultrasound through the surface medium ($\rho_2$ and $v_2$ are defined similarly for the carrier medium).

$$u = 0 \qquad \text{on } \partial\Omega.$$

This case is called soft scatterer.

(iii) **Neumann Boundary Condition**

This condition is imposed for bounded domains when the surface material has much higher acoustic impedance than the host medium ($\rho_2 v_2 \gg \rho_1 v_1$),

$$\frac{\partial u}{\partial \mathbf{n}} = 0 \qquad \text{on } \partial\Omega,$$

where $\mathbf{n}$ is the outer normal vector to $\partial\Omega$. This case is called hard scatterer.

(iv) **Robin Boundary Condition**

This condition *models* the acoustic impedance of the boundary,

$$\frac{\partial u}{\partial \mathbf{n}} + i\sigma u = 0 \qquad \text{on } \partial\Omega,$$

where $\sigma$ is a parameter which measure the admittance of the surface.

### 2.2.2 Helmholtz Model Problem

Model Problem:

Let $\Omega \subset \mathbb{R}^d$ with $d = 2, 3$ be a bounded Lipschitz domain. Also assume the right-hand side $f \in L^2(\Omega)$ and $g \in L^2(\partial\Omega)$,

$$-\Delta u - k^2 u = f \qquad \text{in } \Omega, \tag{2.4a}$$

$$\frac{\partial u}{\partial \mathbf{n}} + iku = g \qquad \text{on } \partial\Omega. \tag{2.4b}$$

[1] The acoustic impedance is a material property which is defined by the density of the medium times the propagation velocity of ultrasound through the medium.

### 2.2.3 The Abstract Variational Formulation

Let $\Omega$ be a bounded Lipschitz domain. The weak formulation of the problem (2.4) is given by:

Find $u \in V := H^1(\Omega)$ such that

$$a(u,v) + b(u,v) = F(v) \qquad \forall v \in H^1(\Omega), \tag{2.5}$$

where

$$a(u,v) := \int_\Omega \left(\nabla u \nabla \bar{v} - k^2 u \bar{v}\right) dV, \qquad \forall u, v \in H^1(\Omega) \tag{2.6}$$

$$b(u,v) := \mathrm{i}k \int_{\partial\Omega} u \bar{v} dS, \qquad \forall u, v \in H^1(\Omega) \tag{2.7}$$

$$F(v) := \int_\Omega f \bar{v} dV + \int_{\partial\Omega} g \bar{v} dS \qquad \forall v \in H^1(\Omega). \tag{2.8}$$

**Proposition 2.1.** *([43, Prop. 8.1.3]) Let $\Omega$ be a bounded Lipschitz domain. Then, (2.4) is uniquely solvable for all $f \in \left(H^1(\Omega)\right)'$, $g \in H^{-1/2}(\partial\Omega)$ and the solution depends continuously on the data.*

*Proof.* For the proof we refer to the Proposition 8.1.3 in [43].

## 2.3 Finite Element Approximations

### 2.3.1 Notation for Function Spaces

In the theory of the elliptic partial differential equations, it is common to work with the Sobolev spaces. We use the standard notations. We denote the Sobolev space $W_p^m(\Omega)$ for $1 \le p \le \infty$ as

$$W_p^m(\Omega) := \{u \in L^p(\Omega) : D^\alpha u \in L^p(\Omega) \quad \forall |\alpha| \le m\},$$

where $\alpha$ is a multi-index and the weak derivative[2] of order $\alpha$ is denoted by $D^\alpha u$. We denote the Hilbert space $H^m(\Omega)$ as a special case of the Sobolev space for $p = 2$, with

[2] Let $u \in L^1(\Omega)$. Then $v$ is the weak derivative of $u$ if and only if

$$\int_\Omega u D^\alpha \phi dx = (-1)^{|\alpha|} \int_\Omega v \phi dx \quad \forall \phi \in C_0^\infty(\Omega),$$

where

$$D^\alpha u = \frac{\partial^{|\alpha|} u}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}.$$

the following seminorm and norm,

$$|u|^2_{H^m(\Omega)} = \sum_{|\alpha|=m} \int_\Omega |D^\alpha u|^2 d\mathbf{x} \quad \text{for } m \in \mathbb{N}_0. \tag{2.9}$$

$$\|u\|^2_{H^m(\Omega)} = \sum_{|\alpha|\leq m} \int_\Omega |D^\alpha u|^2 d\mathbf{x} \quad \text{for } m \in \mathbb{N}_0. \tag{2.10}$$

We equip the space $H^1(\Omega)$ with the norm

$$\|u\|_{\mathcal{H},\Omega} := \left(|u|^2_{H^1(\Omega)} + k^2\|u\|^2_{L^2(\Omega)}\right)^{1/2}, \tag{2.11}$$

which is equivalent to the $H^1(\Omega)$-norm for $k \geq k_0 > 0$.

### 2.3.2 Galerkin Discretization

Let $S_c$ be a finite dimensional subspace of $H^1(\Omega)$. The conforming Galerkin finite element discretization of (2.5) is as follows
Find $u_{S_c} \in S_c$ such that

$$a(u_{S_c}, v) + b(u_{S_c}, v) = F(v) \qquad \forall v \in S_c, \tag{2.12}$$

where

$$a(u,v) = \int_\Omega \left(\nabla u \nabla \bar{v} - k^2 u\bar{v}\right) dV, \qquad \forall u, v \in S_c \tag{2.13}$$

$$b(u,v) = \mathrm{i}k \int_{\partial\Omega} u\bar{v} dS, \qquad \forall u, v \in S_c \tag{2.14}$$

$$F(v) = \int_\Omega f\bar{v} dV + \int_{\partial\Omega} g\bar{v} dS \qquad \forall v \in S_c. \tag{2.15}$$

The results of the following subsections are taken mainly from [46, 47].

### 2.3.3 Stability and Convergence Analysis

**Theorem 2.2.** *([47, Theorem 3.2, Corollary 3.3])*
*Let $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$, be a bounded Lipschitz domain. Let the sesquilinear forms $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$ be given by (2.6) and (2.7). Then we have,*

*(i)* $b : H^1(\Omega) \times H^1(\Omega) \to \mathbb{C}$ *is a continuous sesquilinear form with*

$$|b(u,v)| \leq C_b \|u\|_{\mathcal{H},\Omega} \|v\|_{\mathcal{H},\Omega} \quad \forall u, v \in H^1(\Omega). \tag{2.16}$$

*where the the constant $C_b$ depends solely on $\Omega$.*

*(ii) The following Gårding inequality holds:*

$$\mathrm{Re}(a(u,v)+b(u,v))+2k^2\|u\|^2_{L^2(\Omega)} \geq \|u\|^2_{\mathcal{H},\Omega} \quad \forall u \in H^1(\Omega). \tag{2.17}$$

*(iii) The adjoint problem to (2.5) which is defined by*

*Find $\phi \in H^1(\Omega)$ such that*

$$\int_\Omega \nabla\phi\nabla\bar{\psi}dV - k^2\int_\Omega \phi\bar{\psi}dV + ik\int_{\partial\Omega}\phi\bar{\psi}dS = \int_\Omega w\bar{\psi}dV \qquad \forall\psi \in H^1(\Omega) \tag{2.18}$$

*is uniquely solvable for every $f \in L^2(\Omega)$.*

*Let $N_k^* : f \to \phi$ denote the adjoint solution operator with*

$$C_{adj} := \sup_{f\in L^2(\Omega)\setminus\{0\}} \frac{\|N_k^* f\|_{\mathcal{H},\Omega}}{\|f\|_{L^2(\Omega)}}. \tag{2.19}$$

*Let $S_c$ be a closed subspace of $H^1(\Omega)$ and define the adjoint approximability*

$$\eta(S_c) := \sup_{f\in L^2(\Omega)\setminus\{0\}} \inf_{v\in S_c} \frac{\|N_k^* f - v\|_{\mathcal{H},\Omega}}{\|f\|_{L^2(\Omega)}}. \tag{2.20}$$

*Then, the condition*

$$k\eta(S_c) \leq \frac{1}{4(1+C_b)} \tag{2.21}$$

*implies the following statements:*

*(iv) The discrete inf-sup condition is satisfied*

$$\inf_{u\in S_c\setminus\{0\}} \sup_{v\in S_c\setminus\{0\}} \frac{|a(u,v)+b(u,v)|}{\|u\|_{\mathcal{H},\Omega}\|v\|_{\mathcal{H},\Omega}} \geq \frac{1}{2+1/(1+C_b)+4kC_{adj}} > 0. \tag{2.22}$$

*(iiv) (quasi optimality) For every $u \in \mathcal{H}$ there exists a unique $u_{S_c}$ with Galerkin orthogonality property*

$$a(u-u_{S_c},v)+b(u-u_{S_c},v) = 0 \qquad \forall v \in S_c,$$

*and there holds*

$$\|u-u_{S_c}\|_{\mathcal{H},\Omega} \leq 2(1+C_b)\inf_{v\in S_c}\|u-v\|_{\mathcal{H},\Omega}, \tag{2.23}$$

$$\|u-u_{S_c}\|_{L^2(\Omega)} \leq (1+C_b)\eta(S_c)\inf_{v\in S_c}\|u-u_{S_c}\|_{\mathcal{H},\Omega}, \tag{2.24}$$

*Proof.* For a proof we refer to the Theorem 3.2 and Corollary 3.3 in [47]. □

**Remark 2.3.** The similar results have been proved in [47, 46] for the case of Helmholtz problem with Dirichlet boundary condition and the case of Helmholtz problem with Sommerfeld radiation condition for the exterior domains.

From Theorem 2.2 we conclude that the adjoint approximation property $\eta_{S_c}$ plays the key role for the stability and convergence estimates. This quantity allows to quantify the quality of a (new) approximation space $S_c$ for the Galerkin discretization. In the following, we will present a theory which allows to estimate this quantity.

### 2.3.4 Stable Decomposition

The main idea of the refined regularity results in [46] is based on the frequency splitting of the right-hand side and the estimation of the solution operators applied to the high and low frequency parts of it. This splitting is based on the Fourier transform. The different cases of such an splitting were studied for the Helmholtz problem

(i) with Sommerfeld radiation condition, in [46],

(ii) with Dirichlet boundary condition for an exterior domain, in [47],

(iii) with Robin boundary condition on a bounded Lipschitz domain, in [47].

As our focus in this thesis is on the Helmholtz problem of type (iii), we only recall here the decomposition results for bounded domains.

**Assumption 2.4.** Let $u$ be the solution of the Helmholtz problem. Then it satisfies the following estimate:

$$\|u\|_{\mathcal{H},\Omega} \leq C_S k^{\vartheta} \left( \|f\|_{L^2(\Omega)} + \|g\|_{L^2(\partial\Omega)} \right) \tag{2.25}$$

for some $C_S$ and $\vartheta \geq 0$ independent of $k$.

For the Helmholtz problem with Robin boundary conditions (2.4) the assumption (2.4) is fulfilled with $\vartheta = 5/2$ by [22, Thm. 2.4]. For smooth domains which are star-shaped with respect to a ball or convex polygon, $\vartheta = 0$ is possible as shown in [43, Prop. 8.1.4] for $d = 2$ and subsequently for $d = 3$ in [14].

**Theorem 2.5.** *([47, Theorem 4.10]) (Decomposition for bounded domain).*
*Consider the model problem (2.4). Assume that $\Omega$ is a bounded Lipschitz domain with an analytic boundary or is a convex polygonal domain. Then, there exist constants $C, \lambda > 0, \overrightarrow{\beta} \in [0,1)^J$ independent of $k$ such that for every $f \in L^2(\Omega)$ and $g \in H^{1/2}(\partial\Omega)$*

*the solution $u = S_k(f,g)$ of the Helmholtz problem (2.4) can be written as $u = u_{H^2} + u_{\mathcal{A}}$, where, for all $p \in \mathbb{N}_0$*

$$\|u_{\mathcal{A}}\|_{\mathcal{H},\Omega} \leq Ck^{\vartheta}\left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\right) \tag{2.26}$$

$$\|\Phi_{p,\vec{\beta},k}\nabla^{p+2}u_{\mathcal{A}}\|_{L^2(\Omega)} \leq C\lambda^p k^{\vartheta-1}\max\{p,k\}^{p+2}\left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\right) \tag{2.27}$$

$$\|u_{H^2}\|_{H^2(\Omega)} + k\|u_{H^2}\|_{\mathcal{H},\Omega} \leq C\left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\right), \tag{2.28}$$

*where the weight functions $\Phi_{p,\vec{\beta},k}$ are defined as follows:*

$$\Phi_{p,\beta,k}(x) = \min\left\{1, \frac{|x|}{\min\left\{1, \frac{|p|+1}{k+1}\right\}}\right\}^{p+\beta}, \quad \text{for}\,\beta \in [0,1),\ p \in \mathbb{N}_0,\ \text{and}\ k > 0.$$

**Remark 2.6.** In Theorem 2.5, $\Phi_{p,\vec{\beta},k} = 1$ if $\Omega$ has an analytic boundary while it is the weight function which takes into account corner singularities of the solution for polygonal domains (cf. [47]). Later to avoid weight functions we restrict ourselves most of the time to the case of the domains with analytic boundaries.

*Proof.* (proof of Thm. 2.5) We recall some parts of the proof from [47]. First of all, we note that from the linearity of the operator $S_k$ it is enough to consider the decomposition of $u = S_k(f,0)$ and $u = S_k(0,g)$ separately.

We subdivide this proof into three main steps. The first two steps are related to the properties of $S_k(f,0)$ and $S_k(0,g)$ using a tool to decompose the solution into two parts (i.e. $H^2$ part and analytic part), and in the last step we show how to get such an estimate for the solution of the Helmholtz equation using the previous steps.

**1st step.**(properties of $S_k(f,0)$)

Let $\Omega$ be a bounded Lipschitz domain and $q \in (0,1)$, then there exist constants $C$, $\lambda > 0, \vec{\beta} \in [0,1)^J$ independent of $k$ such that for every $f \in L^2(\Omega)$ the solution $u = S_k(f,0)$ of the Helmholtz problem (2.4) can be written as $u = u_{H^2} + u_{\mathcal{A}} + \tilde{u}$, where, for all $p \in \mathbb{N}_0$

$$\|u_{\mathcal{A}}\|_{\mathcal{H},\Omega} \leq Ck^{\vartheta}\|f\|_{L^2(\Omega)} \tag{2.29}$$

$$\|\Phi_{p,\vec{\beta},k}\nabla^{p+2}u_{\mathcal{A}}\|_{L^2(\Omega)} \leq C\lambda^p k^{\vartheta-1}\max\{p+2,k\}^{p+2}\|f\|_{L^2(\Omega)} \tag{2.30}$$

$$\|u_{H^2}\|_{\mathcal{H},\Omega} \leq qk^{-1}\|f\|_{L^2(\Omega)}, \tag{2.31}$$

$$\|u_{H^2}\|_{H^2(\Omega)} \leq C\|f\|_{L^2(\Omega)}, \tag{2.32}$$

and the remainder $\tilde{u} = S_k(\tilde{f},0)$ satisfies

$$-\Delta\tilde{u} - k^2\tilde{u} = \tilde{f}, \tag{2.33}$$

$$\partial_{\mathbf{n}}\tilde{u} - ik\tilde{u} = 0, \tag{2.34}$$

where

$$\|\tilde{f}\|_{L^2(\Omega)} \leq q\|f\|_{L^2(\Omega)}.$$

**2nd step.** (properties of $S_k(0,g)$)
Let $\Omega$ be a bounded Lipschitz domain and $q \in (0,1)$, then there exist constants $C, \lambda > 0, \overrightarrow{\beta} \in [0,1)^J$ independent of $k$ such that for every $g \in H^{1/2}(\partial\Omega)$ the solution $u = S_k(0,g)$ of the Helmholtz problem (2.4) can be written as $u = u_{H^2} + u_{\mathcal{A}} + \tilde{u}$, where, for all $p \in \mathbb{N}_0$

$$\|u_{\mathcal{A}}\|_{\mathcal{H},\Omega} \leq Ck^{\vartheta}\|g\|_{H^{1/2}(\partial\Omega)} \tag{2.35}$$

$$\|\Phi_{p,\overrightarrow{\beta},k}\nabla^{p+2}u_{\mathcal{A}}\|_{L^2(\Omega)} \leq C\lambda^p k^{\vartheta-1}\max\{p+2,k\}^{p+2}\|g\|_{H^{1/2}(\partial\Omega)} \tag{2.36}$$

$$\|u_{H^2}\|_{H^2(\Omega)} \leq qk^{-1}\|g\|_{H^{1/2}(\partial\Omega)}, \tag{2.37}$$

$$\|u_{H^2}\|_{\mathcal{H},\Omega} \leq C\|g\|_{H^{1/2}(\partial\Omega)}. \tag{2.38}$$

The remainder $\tilde{u} = S_k(0,\tilde{g})$ satisfies for a $\tilde{g}$ with $\|\tilde{g}\|_{H^{1/2}(\partial\Omega)} \leq q\|g\|_{H^{1/2}(\partial\Omega)}$,

$$-\Delta\tilde{u} - k^2\tilde{u} = 0, \tag{2.39}$$

$$\partial_{\mathbf{n}}\tilde{u} - ik\tilde{u} = \tilde{g}. \tag{2.40}$$

**3rd step.**
Let $f^{(0)} := f$, then from part (i) for $u = S_k(f^{(0)},0)$ we get

$$u = u_{\mathcal{A}}^{(0)} + u_{H^2}^{(0)} + S_k(f^{(1)},0), \quad \text{for some} \quad f^{(1)} \in L^2(\Omega),$$

where $u_{\mathcal{A}}^{(0)}$ and $u_{H^2}^{(0)}$ satisfy the estimates stated in (i) and $\|f^{(1)}\|_{L^2(\Omega)} \leq q\|f^{(0)}\|_{L^2(\Omega)}$ for some $q \in (0,1)$. Now we may iterate this argument and we can write $u$ as a sum of series (of analytic functions and $H^2$-functions) which can be bounded by using a geometric series argument. The same argument can be applied to $S_k(0,g)$. □

# 3

# Discontinuous Galerkin Methods

It is known that the use of conforming methods to solve the variational equation (2.12) requires to impose a minimal resolution condition to prove discrete stability. For example $kh/p < c_1$ and $p > c_2 \log k$ with sufficiently small $c_1$ and sufficiently large $c_2$ are necessary conditions to get the stability for a large class of Helmholtz problems (e.g., (i) for bounded domains with analytic boundaries and Robin boundary condition, (ii) for convex two dimensional polygons, (iii) for exterior domains in $\mathbb{R}^d$ ($d \in \{2,3\}$) with analytic boundaries and Dirichlet boundary conditions).

In the context of non-conforming discontinuous Galerkin methods (DG methods) our goal is to derive a discretization of the Helmholtz problem which is unconditionally stable, i.e., the existence and uniqueness of the discrete solution is guaranteed without a resolution condition on the discrete space. In the recent years DG methods have become very popular. The major advantage of DG methods is that "strict" conditions such as, e.g., continuity in the domain or essential boundary conditions or stabilized terms can be "built" weakly into the variational formulation. This typically allows to use and construct appropriate finite element space in a much more flexible way.

In this chapter we will derive a general discontinuous Galerkin method for the Helmholtz problem. By a certain choice of the flux parameters we will derive the *ultra weak variational formulation*.

## 3.1 UWVF for the Helmholtz Problem

In the class of DG methods there is a method which is called the ultra weak variational formulation (UWVF). The UWVF can be regarded a special DG method as discussed in [8, 28]. The ultra weak variational formulation was introduced by B. Després in [16] and it was applied for the Helmholtz problem as well as Maxwell equations in

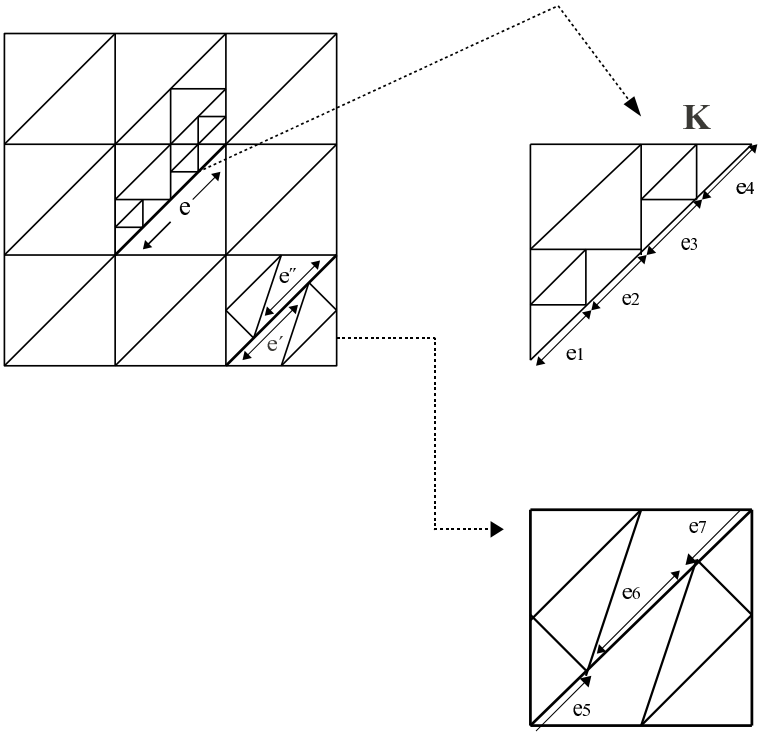


**Figure 3.1**: A possible non-conforming triangulation

[9, 10]. Here we follow the same technique as in [28] to get the ultra weak variational formulation for the Helmholtz problem.

Consider the model problem introduced in Section 2.2.2. The starting point is to write the problem as a first order system. We introduce an auxiliary variable $\sigma := \nabla u/ik$ and insert it into the equations (2.4a) and (2.4b),

$$\begin{aligned} ik\sigma &= \nabla u && \text{in } \Omega, \\ iku - \nabla\cdot\sigma &= \frac{1}{ik} f && \text{in } \Omega, \\ ik\sigma\cdot\mathbf{n} + iku &= g && \text{on } \partial\Omega. \end{aligned} \tag{3.1}$$

Let $\mathcal{T}$ denote a partition of $\Omega$ into non-overlapping polygonal/polyhedral subdomains ("finite elements") $K$ of diameters $h_K$ with possibly hanging nodes[1]. We define

[1] A vertex of an element is called a hanging node if it lies in the interior of an edge or face of another element.

for each element the following sets

$$\mathcal{E}^{\mathcal{I}}(K) := \{E : E \text{ is an (full) interior edge } (d=2), \text{ or an interior face } (d=3) \text{ of } K\}, \tag{3.2}$$

$$\mathcal{E}^{\mathcal{B}}(K) := \{E : E \text{ is a (full) boundary edge } (d=2), \text{ or a boundary face } (d=3) \text{ of } K\}, \tag{3.3}$$

$$\mathcal{E}(K) := \{E : E \text{ is an (full) edge } (d=2), \text{ or a face } (d=3) \text{ of } K\}, \tag{3.4}$$

and the constant

$$d_{\mathcal{T}} := \max\left\{\sharp\mathcal{E}^{\mathcal{I}}(K) : K \in \mathcal{T}\right\} \tag{3.5}$$

for example, $d_{\mathcal{T}} = d+1$ for simplicial meshes.

The set of the inner and boundary edges in $\mathcal{T}$ are as follows:

$$\mathcal{F}_{\mathcal{T}}^{\mathcal{I}} := \{e : e \in \mathcal{E}^{\mathcal{I}}(K) \quad \text{for some } K \in \mathcal{T}\}, \tag{3.6}$$

$$\mathcal{F}_{\mathcal{T}}^{\mathcal{B}} := \{e : e \in \mathcal{E}^{\mathcal{B}}(K) \quad \text{for some } K \in \mathcal{T}\}, \tag{3.7}$$

$$\tag{3.8}$$

As a convention we assume that the finite elements $K$ are open sets as well as the elements in $\mathcal{F}_{\mathcal{T}}^{\mathcal{I}}$ are assumed to be relatively open.

The inner and the boundary skeletons of $\mathcal{T}$ are given by

$$\mathfrak{S}_{\mathcal{T}}^{\mathcal{I}} := \bigcup_{e\in\mathcal{F}_{\mathcal{T}}^{\mathcal{I}}} e \quad \text{and} \quad \mathfrak{S}_{\mathcal{T}}^{\mathcal{B}} := \bigcup_{e\in\mathcal{F}_{\mathcal{T}}^{\mathcal{B}}} e. \tag{3.9}$$

For later use, we define a refined partition $\mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}}$ of $\mathfrak{S}_{\mathcal{T}}^{\mathcal{I}}$ as the smallest set (with largest pieces) with the properties:

a. For all $e \in \mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}}$ and for all $e' \in \mathcal{F}_{\mathcal{T}}^{\mathcal{I}}$, the intersection $e \cap e'$ is either empty or $e$.

b. $\displaystyle\bigcup_{e\in\mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}}} \overline{e} = \overline{\mathfrak{S}_{\mathcal{T}}^{\mathcal{I}}}$.

As a consequence the following holds
a) the elements in $\mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}}$ are disjoint, and
b) for all $e \in \mathcal{F}_{\mathcal{T}}^{\mathcal{I}}$, there exists a set of sons denoted by $\text{sons}(e) \subset \mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}}$ such that

$$\overline{e} = \bigcup_{e'\in\text{sons}(e)} \overline{e'}. \tag{3.10}$$

The set $\mathcal{F}_{\mathcal{T}}^{\mathcal{B},\text{micro}}$ is defined analogously.

For example in Figure 3.1, we can see that the edges $e$, $e'$ and $e''$ do not belong to $\mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}}$ but we can write them as follows

$$\begin{aligned}\overline{e} &= \overline{e_1} \cup \overline{e_2} \cup \overline{e_3} \cup \overline{e_4},\\ \overline{e'} &= \overline{e_5} \cup \overline{e_6},\\ \overline{e''} &= \overline{e_6} \cup \overline{e_7},\end{aligned}$$

where $e_i \in \mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}}$ for $i \in \{1,2,3,4,5,6,7\}$.

We multiply the first equation in (3.1) by a smooth test function $\tau \in H(\text{div};K) := \{u \in L^2(K) : \text{div}(u) \in L^2(K)\}$ and the second equation in (3.1) by another test function $v \in H^1(K)$. We employ elementwise integration by parts for each of those equations. Then (3.1) is equivalent to:
Find $\sigma \in H(\text{div},K)$ and $u \in H^1(K)$ such that

$$\begin{aligned}\int_K ik\sigma\cdot\overline{\tau}dV + \int_K u\overline{\nabla\cdot\tau}dV - \int_{\partial K} u\overline{\tau\cdot\mathbf{n}}dS &= 0 \qquad \forall\tau \in H(\text{div};K) \quad (3.11)\\ \int_K iku\overline{v}dV + \int_K \sigma\cdot\overline{\nabla v}dV - \int_{\partial K} \sigma\cdot\mathbf{n}\overline{v}dS &= \frac{1}{ik}\int_K f\overline{v}dV \qquad \forall v \in H^1(K).\end{aligned}$$

Note that problem (3.11) is formulated at the continuous level. Now we will replace the spaces $H^1(K)$ and $H(\text{div},K)$ by finite-dimensional subsets $S \subset H^1(K)$ and $\Sigma_S \subset H(\text{div},K)$. We emphasize that in the following theory we assume that the finite dimensional space satisfy the following minimal requirements

$$S \subset L^2(\Omega) \quad \text{and} \quad S \subset \prod_{K\in\mathcal{T}} H^2(K), \tag{3.12}$$

but do not assume that $S$ is piecewise polynomial or a classical finite element space. For the error analysis we will impose further (abstract) condition in the form of discrete trace estimates. Additionally , we impose a coupling between neighboring elements by replacing the multi-valued traces $u$ and $\sigma$ on the element edges by the so-called numerical fluxes $\widehat{u}_S$ and $\widehat{\sigma}_S$. To define the numerical fluxes and present the equivalent problem to (3.11) we need the following definitions.

As a consequence of (3.9), every edge $e \in \mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}}$ is contained in the boundaries of exactly two finite elements which we denote by $K_e^+$ and $K_e^-$. The one-sided restrictions of some $\mathcal{T}$-piecewise smooth function $v$ to $e \in \mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}}$ are denoted by

$$v^+(x) := \lim_{\substack{y\in K^+\\ y\to x}} v(y) \quad \text{and} \quad v^-(x) := \lim_{\substack{y\in K^-\\ y\to x}} v(y) \quad \text{for all } x \in e,$$

and we use the same notation for vector-valued functions such as $\sigma_S$.

The averages and jumps are defined on $e \in \mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}}$ by

$$\begin{aligned}
&\text{the averages:} \quad \{v_S\} := \frac{1}{2}(v_S^+ + v_S^-) \quad , \quad \{\sigma_S\} := \frac{1}{2}(\sigma_S^+ + \sigma_S^-),\\
&\text{the jumps:} \quad [\![v_S]\!]_N := v_S^+\mathbf{n}^+ + v_S^-\mathbf{n}^- \quad , \quad [\![\sigma_S]\!]_N := \sigma_S^+.\mathbf{n}^+ + \sigma_S^-.\mathbf{n}^-,
\end{aligned}$$

where $v_S$ and $\sigma_S$ are piecewise smooth function and vector field on $\mathcal{T}$. The general form of the numerical fluxes is (cf. [28]):
On $\partial K^- \cap \partial K^+ \subset \mathfrak{S}_{\mathcal{T}}^{\mathcal{I}}$, define

$$\begin{aligned}
\widehat{\sigma_S} &= \frac{1}{ik}\{\nabla_{\mathcal{T}} u_S\} - \alpha[\![u_S]\!]_N - \frac{\gamma}{ik}[\![\nabla_{\mathcal{T}} u_S]\!]_N,\\
\widehat{u_S} &= \{u_S\} + \gamma \cdot [\![u_S]\!]_N - \frac{\beta}{ik}[\![\nabla_{\mathcal{T}} u_S]\!]_N,
\end{aligned} \tag{3.13}$$

and on $\partial K \cap \partial\Omega \subset \mathfrak{S}_{\mathcal{T}}^{\mathcal{B}}$, define

$$\begin{aligned}
\widehat{\sigma_S} &= \frac{1}{ik}\nabla_{\mathcal{T}} u_S - (1-\delta)\left(\frac{1}{ik}\nabla_{\mathcal{T}} u_S + u_S\mathbf{n} - \frac{1}{ik}g\mathbf{n}\right),\\
\widehat{u_S} &= u_S - \delta\left(\frac{1}{ik}\nabla_{\mathcal{T}} u_S \cdot \mathbf{n} + u_S - \frac{1}{ik}g\right),
\end{aligned} \tag{3.14}$$

with parameters $\alpha > 0$, $\beta \geq 0$, $\gamma$ and $0 < \delta < 1$. By $\nabla_{\mathcal{T}}$[2] and $\Delta_{\mathcal{T}}$ [3] we denote the elementwise applications of the operators $\nabla$ and $\Delta$, respectively. For the method by Cessenat and Després [9] the parameters were chosen by

$$\alpha = 1/2, \quad \beta = 1/2, \quad \gamma = 0, \quad \delta = 1/2,$$

and the method in [28] is recovered by

$$\alpha = a/kh, \quad \beta = bkh, \quad \gamma = 0, \quad \delta = dkh,$$

where the local mesh size function $h$ is defined on $\mathcal{F}_{\mathcal{T}}^{\mathcal{I}}$ by

$$h(x) = \min\{h_{K^-}, h_{K^+}\}$$

if $x$ is in the interior of $\partial K^- \cap \partial K^+$, where $h_K := \text{diam}K$. For our purpose we take $\gamma = 0$ and determine the suitable choices of $\alpha$,$\beta$ and $\gamma$ later in Chapter 4.

[2] For $v \in H^1(\Omega,\mathcal{T})$ we define $(\nabla_{\mathcal{T}} v)|_K = \nabla(v|_K)$.
[3] For $v \in H^1(\Omega,\mathcal{T})$ we define $(\Delta_{\mathcal{T}} v)|_K = \Delta(v|_K)$.

Now we introduce the discrete solutions $\sigma_S \in \Sigma_S$ and $u_S \in S$ as the solutions of the following system:

$$\begin{aligned}
&\int_K ik\sigma_S \cdot \overline{\tau}_S dV + \int_K u_S \overline{\nabla \cdot \tau_S} dV - \int_{\partial K} \widehat{u}_S \overline{\tau_S \cdot \mathbf{n}} dS = 0 && \forall \tau_S \in \Sigma_S(K),\\
&\int_K iku_S \overline{v}_S dV + \int_K \sigma_S \cdot \overline{\nabla v_S} dV - \int_{\partial K} \widehat{\sigma}_S \cdot \mathbf{n} \overline{v}_S dS = \frac{1}{ik} \int_K f \overline{v}_S dV && \forall v_S \in S(K).
\end{aligned} \tag{3.15}$$

By imposing the assumption $\nabla_{\mathcal{T}} S \subseteq \Sigma_S$ and taking $\tau_S = \nabla v_S$, we can eliminate $\Sigma_S$. We integrate by parts from the first equation in (3.15) and insert the result into the second equation. This results in

$$\int_K (\nabla u_S \cdot \overline{\nabla v_S} - k^2 u_S \overline{v}_S) dV - \int_{\partial K} (u_S - \widehat{u}_S) \overline{\nabla v_S \cdot \mathbf{n}} dS - \int_{\partial K} ik\widehat{\sigma}_S \cdot \mathbf{n} \overline{v}_S dS = \int_K f \overline{v}_S dV. \tag{3.16}$$

By inserting the definitions of the numerical fluxes as in (3.13) and (3.14) into equation (3.16) and summing over all elements, it follows

$$\begin{aligned}
&(\nabla_{\mathcal{T}} u_S, \nabla_{\mathcal{T}} v_S)_{L^2(\Omega)} - k^2 (u_S, v_S)_{L^2(\Omega)} - \int_{\Xi_{\mathcal{T}}^I} [\![ u_S ]\!]_N \cdot \{\overline{\nabla_{\mathcal{T}} v_S}\} dS - \int_{\Xi_{\mathcal{T}}^I} \{\nabla_{\mathcal{T}} u_S\} \cdot [\![ \overline{v}_S ]\!]_N dS \\
&- \int_{\Xi_{\mathcal{T}}^B} \delta u_S \overline{\nabla_{\mathcal{T}} v_S \cdot \mathbf{n}} dS - \int_{\Xi_{\mathcal{T}}^B} \delta \nabla_{\mathcal{T}} u_S \cdot \mathbf{n} \overline{v}_S \, dS - \frac{1}{ik} \int_{\Xi_{\mathcal{T}}^I} \beta [\![ \nabla_{\mathcal{T}} u_S ]\!]_N [\![ \overline{\nabla_{\mathcal{T}} v_S} ]\!]_N dS \\
&- \frac{1}{ik} \int_{\Xi_{\mathcal{T}}^B} \delta \nabla_{\mathcal{T}} u_S \cdot \mathbf{n} \overline{\nabla_{\mathcal{T}} v_S \cdot \mathbf{n}} dS + ik \int_{\Xi_{\mathcal{T}}^I} \alpha [\![ u_S ]\!]_N \cdot [\![ \overline{v}_S ]\!]_N dS + ik \int_{\Xi_{\mathcal{T}}^{\mathcal{B}}} (1-\delta) u_S \overline{v}_S \, dS \\
&+ \int_{\Xi_{\mathcal{T}}^B} \delta \frac{1}{ik} g \overline{\nabla_{\mathcal{T}} v_S \cdot \mathbf{n}} dS - \int_{\Xi_{\mathcal{T}}^B} (1-\delta) g \overline{v}_S \, dS = \quad (f, v_S)_{L^2(\Omega)},
\end{aligned}$$

where we used the following relations

$$\begin{aligned}
\sum_{K \in \mathcal{T}} \int_{\partial K \setminus \partial\Omega} \{\nabla_{\mathcal{T}} u_S\} \cdot \mathbf{n} \overline{v}_S \, dS &= \int_{\Xi_{\mathcal{T}}^I} \{\nabla_{\mathcal{T}} u_S\} \cdot [\![ \overline{v}_S ]\!]_N dS \\
\sum_{K \in \mathcal{T}} \int_{\partial K \setminus \partial\Omega} (u_S - \{u_S\}) \overline{\nabla_{\mathcal{T}} v_S \cdot \mathbf{n}} dS &= \int_{\Xi_{\mathcal{T}}^I} [\![ u_S ]\!]_N \cdot \{\overline{\nabla_{\mathcal{T}} v_S}\} dS.
\end{aligned}$$

Finally the UWVF can be written in the form:

Find $u_S \in S$ such that, for all $v \in S$,

$$a_{\mathcal{T}}(u_S, v) - k^2 (u_S, v)_{L^2(\Omega)} = (f, v)_{L^2(\Omega)} - \int_{\Xi_{\mathcal{T}}^B} \frac{1}{ik} \delta g \overline{\nabla_{\mathcal{T}} v \cdot \mathbf{n}} dS + \int_{\Xi_{\mathcal{T}}^B} (1-\delta) g \overline{v} dS, \tag{3.17}$$

where $a_{\mathcal{T}}(\cdot,\cdot)$ is the DG-sesquilinear form on $S \times S$ defined by

$$\begin{aligned} a_{\mathcal{T}}(u,v) := (\nabla_{\mathcal{T}} u, \nabla_{\mathcal{T}} v)_{L^2(\Omega)} &- \int_{\Xi_{\mathcal{T}}^{\mathcal{I}}} [\![u]\!]_N \cdot \{\overline{\nabla_{\mathcal{T}} v}\}\, dS - \int_{\Xi_{\mathcal{T}}^{\mathcal{I}}} \{\nabla_{\mathcal{T}} u\} \cdot [\![\overline{v}]\!]_N\, dS \\ &- \int_{\Xi_{\mathcal{T}}^{B}} \delta u \overline{\nabla_{\mathcal{T}} v \cdot \mathbf{n}}\, dS - \int_{\Xi_{\mathcal{T}}^{B}} \delta \nabla_{\mathcal{T}} u \cdot \mathbf{n} \overline{v}\, dS \\ &- \frac{1}{ik} \int_{\Xi_{\mathcal{T}}^{\mathcal{I}}} \beta [\![\nabla_{\mathcal{T}} u]\!]_N [\![\overline{\nabla_{\mathcal{T}} v}]\!]_N\, dS - \frac{1}{ik} \int_{\Xi_{\mathcal{T}}^{B}} \delta \nabla_{\mathcal{T}} u \cdot \mathbf{n} \overline{\nabla_{\mathcal{T}} v \cdot \mathbf{n}}\, dS \\ &+ ik \int_{\Xi_{\mathcal{T}}^{\mathcal{I}}} \alpha [\![u]\!]_N \cdot [\![\overline{v}]\!]_N\, dS + ik \int_{\Xi_{\mathcal{T}}^{\mathcal{B}}} (1-\delta) u \overline{v}\, dS. \end{aligned} \tag{3.18}$$

**Remark 3.1.** For functions $u, v \in H^{3/2+\varepsilon}(\Omega)$ all terms in the bilinear form (3.18) are well-defined due to well known mapping properties of the trace and the normal trace operator for Lipschitz domains [17, 29, 56].

# 4

# Stability and Convergence Analysis

## 4.1 Notations

For our stability and convergence estimates we will use mesh dependent norms $\|\cdot\|_{DG}$ and $\|\cdot\|_{DG^+}$ as follows

$$\begin{aligned}\|v\|_{DG}^2 :=& \|\nabla_{\mathcal{T}} v\|_{L^2(\Omega)}^2 + k^{-1}\|\beta^{1/2}[\![\nabla_{\mathcal{T}} v]\!]_N\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)}^2 + k\|\alpha^{1/2}[\![v]\!]_N\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)}^2 \\ &+ k^{-1}\|\delta^{1/2}\nabla_{\mathcal{T}} v\cdot\mathbf{n}\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{B}}\right)}^2 + k\|(1-\delta)^{1/2} v\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{B}}\right)}^2 + k^2\|v\|_{L^2(\Omega)}^2, \\ \|v\|_{DG^+}^2 :=& \|v\|_{DG}^2 + k^{-1}\|\alpha^{-1/2}\{\nabla_{\mathcal{T}} v\}\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)}^2.\end{aligned}$$

Note that these norms are well defined for all $v \in H^{3/2+\varepsilon}(\Omega) + S$ for $\varepsilon > 0$ because of the jump term of the gradient of $v$.

## 4.2 Continuity and Coercivity

We define the auxiliary bilinear form $b_{\mathcal{T}}$ which is related to the operator $-\Delta + k^2$ as follows:

$$b_{\mathcal{T}}(u,v) := a_{\mathcal{T}}(u,v) + k^2(u,v)_{L^2(\Omega)}. \tag{4.1}$$

We prove that this bilinear form is coercive and continuous with respect to the appropriate mesh dependent norm. First we introduce the condition on the flux parameter $\alpha$ which is necessary for our continuity and coercivity analysis.

**Definition 4.1.** *We define the trace constant $C_{\text{trace}}(S,K)$ as the minimal constant in*

$$\max_{e\in\mathcal{E}(K)} \|\nabla(v|_K)\|_{L^2(e)} \le C_{\text{trace}}(S,K)\|\nabla v\|_{L^2(K)} \qquad \forall v \in S, \tag{4.2}$$

*and we denote by $\mathcal{E}(K)$ the set of all edges of $K$.*

**Remark 4.2.** In the definition of the sesquilinear form $a_{\mathcal{T}}(u,v)$ we denote by $\alpha \in L^{\infty}\left(\mathfrak{S}_{\mathcal{T}}^{\mathcal{I}},\mathbb{R}\right)$ a $\mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}}$-piecewise constant, positive function. For $e \in \mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}}$, we write $\alpha_e := \alpha|_e$. The analysis of the continuity and coercivity will lead to the condition

$$\alpha_e > \frac{4d_{\mathcal{T}}}{3k} \max_{\iota\in\{+,-\}} C_{\text{trace}}^2(S,K_e^{\iota}) \qquad \forall e \in \mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}}. \tag{4.3}$$

For the special case that $S$ is a conforming/ nonconforming $hp$-finite element space, the estimate of the approximation property of $S$ with respect to the $\|\cdot\|_{DG}$, respectively $\|\cdot\|_{DG^+}$ norm, (cf. Section 6.2), leads to the choices

$$\alpha_e = O\left(\max_{\iota\in\{+,-\}} \frac{p^2}{kh_{K_e^{\iota}}}\right), \quad \beta = O(\frac{kh}{p}), \quad \delta = \quad O(\frac{kh}{p}), \quad \forall e \in \mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}} \tag{4.4}$$

where the mesh width $h := \max_{K\in\mathcal{T}} h_K$ with $h_K := \text{diam}(K)$ and $p$ denotes the polynomial degree. We used the notation $A = O(B)$ if there are positive constants $c, C$ independent of $A$ and $B$ such that $cB \le A \le CB$.

**Proposition 4.3.** *Let $0 < \delta < 1/3$ and $\alpha$ satisfy*

$$\alpha_e > \frac{4d_{\mathcal{T}}}{3k} \max_{\iota\in\{+,-\}} C_{\text{trace}}^2(S,K_e^{\iota}) \qquad \forall e \in \mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}}, \tag{4.5}$$

*where $d_{\mathcal{T}}$ is defined in equation (3.5). Then, there exist constants $c_{\text{coer}}$, $C_{\text{c}} > 0$ independent of $h$, $k$, $\alpha$, $\beta$, $\delta$, and $C_{\text{trace}}(S,K)$ such that*

**a)** *the sesquilinear form $b_{\mathcal{T}}(\cdot,\cdot)$ is coercive*

$$|b_{\mathcal{T}}(v,v)| \ge c_{\text{coer}}\|v\|_{DG}^2 \qquad \forall v \in S, \tag{4.6}$$

**b)** *and continuous*

$$\begin{aligned} |b_{\mathcal{T}}(v,w)| &\le C_{\text{c}}\|v\|_{DG^+}\|w\|_{DG^+} \qquad \forall v,w \in H^{3/2+\epsilon}+S, \\ |b_{\mathcal{T}}(v,w_S)| &\le C_{\text{c}}\|v\|_{DG^+}\|w_S\|_{DG} \qquad \forall v \in H^{3/2+\epsilon}+S, \quad \forall w_S \in S. \end{aligned} \tag{4.7}$$

*Proof.* a) The definition of $b_{\mathcal{T}}(.,.)$ leads to

$$\begin{aligned} b_{\mathcal{T}}(v,v) = \|\nabla_{\mathcal{T}} v\|_{L^2(\Omega)}^2 &\overbrace{- 2\,\text{Re}\left(\int_{\mathfrak{S}_{\mathcal{T}}^{\mathcal{I}}} [\![v]\!]_N \cdot \{\overline{\nabla_{\mathcal{T}} v}\} dS\right)}^{(i)} \overbrace{- 2\,\text{Re}\left(\int_{\mathfrak{S}_{\mathcal{T}}^{\mathcal{B}}} \delta v \overline{\nabla_{\mathcal{T}} v \cdot \mathbf{n}} dS\right)}^{(ii)} \\ &+ ik^{-1}\|\beta^{1/2}[\![\nabla_{\mathcal{T}} v]\!]_N\|_{L^2\left(\mathfrak{S}_{\mathcal{T}}^{\mathcal{I}}\right)}^2 + ik^{-1}\|\delta^{1/2}\nabla_{\mathcal{T}} v \cdot \mathbf{n}\|_{L^2\left(\mathfrak{S}_{\mathcal{T}}^{\mathcal{B}}\right)}^2 \\ &+ ik\|\alpha^{1/2}[\![v]\!]_N\|_{L^2\left(\mathfrak{S}_{\mathcal{T}}^{\mathcal{I}}\right)}^2 + ik\|(1-\delta)^{1/2} v\|_{0,\mathfrak{S}_{\mathcal{T}}^{\mathcal{B}}}^2 + k^2\|v\|_{L^2(\Omega)}^2. \end{aligned}$$

By using Young's inequality for a positive $\mathcal{F}_{\mathcal{T}}^{\mathcal{I},\mathrm{micro}}$-piecewise constant function $s > 0$ we get for the second term in the representation of $b_{\mathcal{T}}(\cdot,\cdot)$, i.e. for (i),

$$\begin{aligned}\left|2\,\mathrm{Re}\int_{\Xi_{\mathcal{T}}^{\mathcal{I}}}[\![v]\!]_N\cdot\{\overline{\nabla_{\mathcal{T}}v}\}dS\right| &= \left|2\,\mathrm{Re}\sum_{e\in\mathcal{F}_{\mathcal{T}}^{\mathcal{I},\mathrm{micro}}}\int_e[\![v]\!]_N\cdot\{\overline{\nabla_{\mathcal{T}}v}\}dS\right| \\ &\le k\left\|\sqrt{\frac{s}{\alpha}}\alpha^{1/2}[\![v]\!]_N\right\|^2_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)} \\ &\quad+\frac{1}{k}\sum_{K\in\mathcal{T}}\sum_{e\in\mathcal{E}(K)}\sum_{e'\in\mathrm{sons}(e)}\left\|\frac{1}{\sqrt{s}}\nabla(v|_K)\right\|^2_{L^2(e')}.\end{aligned}$$

We choose $s|_{e'} := 4\alpha_{e'}/5$ for $e'\in\mathcal{F}_{\mathcal{T}}^{\mathcal{I},\mathrm{micro}}$ and by using (4.3) we get

$$\begin{aligned}\left|2\,\mathrm{Re}\int_{\Xi_{\mathcal{T}}^{\mathcal{I}}}[\![v]\!]_N\cdot\{\overline{\nabla_{\mathcal{T}}v}\}dS\right| &\le \frac{4}{5}k\left\|\alpha^{1/2}[\![v]\!]_N\right\|^2_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)} \\ &\quad+\sum_{K\in\mathcal{T}}\sum_{e\in\mathcal{E}^{\mathcal{I}}(K)}\frac{15}{16d_{\mathcal{T}}}\sum_{e'\in\mathrm{sons}(e)}\frac{1}{\max_{\iota\in\{+,-\}}C^2_{\mathrm{trace}}(S,K^{\iota}_{e'})}\|\nabla(v|_K)\|^2_{L^2(e')}.\end{aligned}$$

Note that $K\in\{K^+_{e'},K^-_{e'}\}$ so that $C_{\mathrm{trace}}(S,K)\le\max_{\iota\in\{+,-\}}C_{\mathrm{trace}}(S,K^{\iota}_{e'})$, so it follows

$$\begin{aligned}\left|2\,\mathrm{Re}\int_{\Xi_{\mathcal{T}}^{\mathcal{I}}}[\![v]\!]_N\cdot\{\overline{\nabla_{\mathcal{T}}v}\}dS\right| &\le \frac{4}{5}k\left\|\alpha^{1/2}[\![v]\!]_N\right\|^2_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)} \\ &\quad+\sum_{K\in\mathcal{T}}\sum_{e\in\mathcal{E}^{\mathcal{I}}(K)}\frac{15}{16d_{\mathcal{T}}C^2_{\mathrm{trace}}(S,K)}\sum_{e'\in\mathrm{sons}(e)}\|\nabla(v|_K)\|^2_{L^2(e')}.\end{aligned}$$

For the second summand we get

$$\begin{aligned}&\sum_{K\in\mathcal{T}}\sum_{e\in\mathcal{E}^{\mathcal{I}}(K)}\frac{15}{16d_{\mathcal{T}}}\frac{1}{C^2_{\mathrm{trace}}(S,K)}\sum_{e'\in\mathrm{sons}(e)}\|\nabla(v|_K)\|^2_{L^2(e')} \\ &\qquad=\sum_{K\in\mathcal{T}}\sum_{e\in\mathcal{E}^{\mathcal{I}}(K)}\frac{15}{16d_{\mathcal{T}}}\frac{1}{C^2_{\mathrm{trace}}(S,K)}\|\nabla(v|_K)\|^2_{L^2(e)} \\ &\qquad\overset{(4.2)}{\le}\frac{15}{16d_{\mathcal{T}}}\sum_{K\in\mathcal{T}}\sum_{e\in\mathcal{E}^{\mathcal{I}}(K)}\|\nabla v\|^2_{L^2(K)} \\ &\qquad\le\frac{15}{16}\|\nabla_{\mathcal{T}}v\|^2_{L^2(\Omega)}.\end{aligned}$$

All in all we have derived

$$\left|2\,\mathrm{Re}\int_{\Xi_{\mathcal{T}}^{\mathcal{I}}}[\![v]\!]_N\cdot\{\overline{\nabla_{\mathcal{T}}v}\}dS\right|\leq\frac{4k}{5}\|\alpha^{1/2}[\![v]\!]_N\|^2_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)}+\frac{15}{16}\|\nabla_{\mathcal{T}}v\|^2_{L^2(\Omega)}.$$

The third term in $b_{\mathcal{T}}(\cdot,\cdot)$, i.e. (ii), can be estimated in a similar fashion for any $t>0$ by

$$\left|2\,\mathrm{Re}\int_{\Xi_{\mathcal{T}}^{\mathcal{B}}}\delta v\overline{\nabla_{\mathcal{T}}v}\cdot\mathbf{n}dS\right|\leq tk\frac{\delta}{1-\delta}\|(1-\delta)^{1/2}v\|^2_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{B}}\right)}+\frac{1}{tk}\|\delta^{1/2}\nabla_{\mathcal{T}}v\cdot\mathbf{n}\|^2_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{B}}\right)}.$$

By choosing $t=3/2$ and $0<\delta<1/3$ we obtain

$$\begin{aligned}
|b_{\mathcal{T}}(v,v)|&\geq\frac{1}{\sqrt{2}}\left(|\mathrm{Re}(b_{\mathcal{T}}(v,v))|+|\mathrm{Im}(b_{\mathcal{T}}(v,v))|\right)\\
&\geq\frac{1}{\sqrt{2}}\left(\frac{1}{16}\|\nabla_{\mathcal{T}}v\|^2_{L^2(\Omega)}+\frac{k}{5}\|\alpha^{1/2}[\![v]\!]_N\|^2_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)}\right.\\
&\quad+\frac{k}{4}\|(1-\delta)^{1/2}v\|^2_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{B}}\right)}+\frac{1}{3k}\|\delta^{1/2}\nabla_{\mathcal{T}}v\cdot\mathbf{n}\|^2_{0,\Xi_{\mathcal{T}}^{\mathcal{B}}}\\
&\quad\left.+k^{-1}\|\beta^{1/2}[\![\nabla_{\mathcal{T}}v]\!]_N\|^2_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)}+k^2\|v\|^2_{L^2(\Omega)}\right)\\
&\geq c_{\mathrm{coer}}\|v\|^2_{DG}.
\end{aligned}$$

b) Using the definition of $b_{\mathcal{T}}$ and applying the triangle inequality as well as Cauchy-

Schwarz inequality with appropriate weights we get

$$
\begin{aligned}
|b_{\mathcal{T}}(v,w)| \leq & |(\nabla_{\mathcal{T}} v, \nabla_{\mathcal{T}} w)| + |\int_{\Xi_{\mathcal{T}}^{\mathcal{I}}} [\![v]\!]_N \cdot \{\overline{\nabla_{\mathcal{T}} w}\} dS| \\
& + |\int_{\Xi_{\mathcal{T}}^{\mathcal{I}}} \{\nabla_{\mathcal{T}} v\} \cdot [\![\overline{w}]\!]_N dS| + |\int_{\Xi_{\mathcal{T}}^{\mathcal{B}}} \delta v \overline{\nabla_{\mathcal{T}} w \cdot \mathbf{n}} dS| \\
& + |\int_{\Xi_{\mathcal{T}}^{\mathcal{B}}} \delta \nabla_{\mathcal{T}} v \cdot \mathbf{n} \overline{w} dS| + \frac{1}{k} |\int_{\Xi_{\mathcal{T}}^{\mathcal{I}}} \left(\beta [\![\nabla_{\mathcal{T}} v]\!]_N [\![\overline{\nabla_{\mathcal{T}} w}]\!]_N\right) dS| \\
& + \frac{1}{k} |\int_{\Xi_{\mathcal{T}}^{\mathcal{B}}} \left(\delta \nabla_{\mathcal{T}} v \cdot \mathbf{n} \overline{\nabla_{\mathcal{T}} w \cdot \mathbf{n}}\right) dS| + |\int_{\Xi_{\mathcal{T}}^{\mathcal{I}}} (k\alpha [\![v]\!]_N [\![\overline{w}]\!]_N) dS| \\
& + k |\int_{\Xi_{\mathcal{T}}^{\mathcal{B}}} (1-\delta) v \overline{w} dS| + k^2 |(v,w)| \\
\leq & \|\nabla_{\mathcal{T}} v\|_{L^2(\Omega)} \|\nabla_{\mathcal{T}} w\|_{L^2(\Omega)} + \|\alpha^{1/2} [\![v]\!]_N\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)} \|\alpha^{-1/2} \{\nabla_{\mathcal{T}} w\}\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)} \\
& + \|\alpha^{-1/2} \{\nabla_{\mathcal{T}} v\}\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)} \|\alpha^{1/2} [\![w]\!]_N\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)} + \|\delta^{1/2} v\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{B}}\right)} \\
& \times \|\delta^{1/2} \nabla_{\mathcal{T}} w \cdot \mathbf{n}\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{B}}\right)} + \|\delta^{1/2} \nabla_{\mathcal{T}} v \cdot \mathbf{n}\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{B}}\right)} \|\delta^{1/2} w\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{B}}\right)} \\
& + k^{-1} \|\beta^{1/2} [\![\nabla_{\mathcal{T}} v]\!]_N\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)} \|\beta^{1/2} [\![\nabla_{\mathcal{T}} w]\!]_N\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)} \\
& + k^{-1} \|\delta^{1/2} \nabla_{\mathcal{T}} v \cdot \mathbf{n}\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{B}}\right)} \|\delta^{1/2} \nabla_{\mathcal{T}} w \cdot \mathbf{n}\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{B}}\right)} + k \|\alpha^{1/2} [\![v]\!]_N\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)} \\
& \times \|\alpha^{1/2} [\![w]\!]_N\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)} + k \|(1-\delta)^{1/2} v\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{B}}\right)} \|(1-\delta)^{1/2} w\|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{B}}\right)} \\
& + k^2 \|v\|_{L^2(\Omega)} \|w\|_{L^2(\Omega)}. \qquad (4.8)
\end{aligned}
$$

For $0 < \delta < 1/2$ and for any $v, w \in H^{3/2+\epsilon} + S$ we finally obtain

$$
|b_{\mathcal{T}}(v,w)| \leq C_c \|v\|_{DG^+} \|w\|_{DG^+}. \qquad (4.9)
$$

Estimates in weaker norms are possible if one of these two functions is purely a finite element function, e.g., $w \in S$. A careful inspection of equation (4.8) shows that the only term which requires the $DG^+$-norm instead of $DG$-norm for $w$ in the continuity

estimate is $\int_{\Xi_{\mathcal{T}}^{\mathcal{I}}} [\![v]\!]_N \cdot \{\overline{\nabla_{\mathcal{T}} w}\} dS$. Using Young's inequality we get

$$\begin{aligned}
\left| \int_{\Xi_{\mathcal{T}}^{\mathcal{I}}} [\![v]\!]_N \cdot \{\overline{\nabla_{\mathcal{T}} w}\} dS \right| &\leq \sum_{e \in \mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}}} \left( k^{1/2} \| [\![v]\!]_N \|_{L^2(e)} \times k^{-1/2} \| \{\overline{\nabla_{\mathcal{T}} w}\} \|_{L^2(e)} \right) \\
&\leq \sum_{K \in \mathcal{T}} \sum_{e \in \mathcal{E}^{\mathcal{I}}(K)} \sum_{e' \in \text{sons}(e)} k^{1/2} \| [\![v]\!]_N \|_{L^2(e')} \times k^{-1/2} \| \{\overline{\nabla_{\mathcal{T}} w}\} \|_{L^2(e')} \\
&\leq \sum_{K \in \mathcal{T}} \left\{ \left( \sum_{e \in \mathcal{E}^{\mathcal{I}}(K)} \sum_{e' \in \text{sons}(e)} k \| [\![v]\!]_N \|^2_{L^2(e')} \right)^{1/2} \right. \\
&\qquad \left. \times \left( \sum_{e \in \mathcal{E}^{\mathcal{I}}(K)} k^{-1} \| \nabla (w|_K) \|^2_{L^2(e)} \right)^{1/2} \right\}. \qquad (4.10)
\end{aligned}$$

We apply the trace inequality as in (4.2) and also use the condition (4.3) to obtain

$$\begin{aligned}
\left| \int_{\Xi_{\mathcal{T}}^{\mathcal{I}}} [\![v]\!]_N \cdot \{\overline{\nabla_{\mathcal{T}} w}\} dS \right| &\leq \sum_{K \in \mathcal{T}} \left\{ \left( \sum_{e \in \mathcal{E}^{\mathcal{I}}(K)} \sum_{e' \in \text{sons}(e)} \frac{3\alpha k^2 \| [\![v]\!]_N \|^2_{L^2(e')}}{4 d_{\mathcal{T}} \max_{\iota \in \{+,-\}} C^2_{\text{trace}}(S, K^{\iota}_{e'})} \right)^{1/2} \right. \\
&\qquad \left. \times \left( \frac{d_{\mathcal{T}}}{k} C^2_{\text{trace}}(S,K) \| \nabla_{\mathcal{T}} w \|^2_{L^2(K)} \right)^{1/2} \right\}. \qquad (4.11)
\end{aligned}$$

Note that $K \in \{K^+_{e'}, K^-_{e'}\}$, so that $C_{\text{trace}}(S,K) \leq \max_{\iota \in \{+,-\}} C_{\text{trace}}(S, K^{\iota}_{e'})$. It follows

$$\begin{aligned}
\left| \int_{\Xi_{\mathcal{T}}^{\mathcal{I}}} [\![v]\!]_N \cdot \{\overline{\nabla_{\mathcal{T}} w}\} dS \right| &\leq \sqrt{\frac{3k}{4 d_{\mathcal{T}}}} \sum_{K \in \mathcal{T}} \left\{ \left( \sum_{e \in \mathcal{E}^{\mathcal{I}}(K)} \| \alpha^{1/2} [\![v]\!]_N \|^2_{L^2(e)} \right)^{1/2} \right. \\
&\qquad \left. \times \left( d_{\mathcal{T}} \| \nabla_{\mathcal{T}} w \|^2_{L^2(K)} \right)^{1/2} \right\} \\
&\leq \sqrt{\frac{3}{2}} k^{1/2} \| \alpha^{1/2} [\![v]\!]_N \|_{L^2\left(\Xi_{\mathcal{T}}^{\mathcal{I}}\right)} \| \nabla_{\mathcal{T}} w \|_{L^2(\Omega)}.
\end{aligned}$$

Hence, we finally get the following continuity result

$$|b_{\mathcal{T}}(v, w_S)| \leq C_c \|v\|_{DG^+} \|w_S\|_{DG} \quad \forall v \in H^{3/2+\epsilon} + S \quad \forall w \in S.$$

Note that the analogue estimate can be derived if the first component is in $S$. □

## 4.3 Discrete Stability

In this section, we will prove that under a very mild condition the UWVF always admits a unique solution in the general discrete space $S$. This is in contrast to conventional Galerkin methods applied to (2.5), where a minimal resolution condition on the finite element space, e.g., on the maximal mesh width, has to be imposed in order to guarantee unique solvability of the discrete equations.

**Theorem 4.4.** *Let the discrete space $S$ satisfy (3.12). Let $\beta \geq 0$, $0 < \delta < 1/3$, and choose $\alpha$ such that (4.3) is satisfied. Then, the UWVF problem (3.17) has a unique solution if*

$$C_S \leq \frac{k}{2(1+C_c)} \quad \textit{with} \quad C_S := \sup_{w_S \in S \cap H_0^2(\Omega)\setminus\{0\}} \inf_{v_S \in S} \frac{\| \langle x, \nabla w_S \rangle - v_S \|_{DG^+}}{\| w_S \|_{L^2(\Omega)}}. \tag{4.12}$$

*Proof.* As is well known the existence of the Galerkin solution is equivalent to the statement

$$\forall w_S \in S \setminus \{0\} \quad \exists v_S \in S \quad \text{s.t.} \quad |a_{\mathcal{T}}(w_S, v_S) - k^2 (w_S, v_S)_{L^2(\Omega)}| > 0. \tag{4.13}$$

We prove this indirectly, by showing the following implication:
For any $w_S \in S$ it holds:

$$\Big(\forall v_S \in S \quad a_{\mathcal{T}}(w_S, v_S) - k^2 (w_S, v_S)_{L^2(\Omega)} = 0\Big) \Rightarrow w_S = 0. \tag{4.14}$$

Our assumption in (4.14) implies for any $w_S \in S$

$$\begin{gathered} \operatorname{Im}\Big(a_{\mathcal{T}}(w_S, v_S) - k^2 (w_S, v_S)_{L^2(\Omega)}\Big) = 0 \\ \text{and} \\ \operatorname{Re}\Big(a_{\mathcal{T}}(w_S, v_S) - k^2 (w_S, v_S)_{L^2(\Omega)}\Big) = 0. \end{gathered} \tag{4.15}$$

First we choose the test function $w_S = v_S$ in (4.15). From the equation for the imaginary part we obtain

$$\begin{aligned} k^{-1} \| \beta^{1/2} [\![ \nabla_{\mathcal{T}} w_S ]\!]_N \|^2_{L^2(\Xi^I_{\mathcal{T}})} + k^{-1} \| \delta^{1/2} \nabla_{\mathcal{T}} w_S \cdot \mathbf{n} \|^2_{L^2(\Xi^{\mathcal{B}}_{\mathcal{T}})} + k \| \alpha^{1/2} [\![ w_S ]\!]_N \|^2_{L^2(\Xi^I_{\mathcal{T}})} \\ + k \| (1-\delta)^{1/2} w_S \|^2_{L^2(\Xi^{\mathcal{B}}_{\mathcal{T}})} = 0, \end{aligned} \tag{4.16}$$

and from the equation for the real part we get

$$\begin{aligned} (\nabla_{\mathcal{T}} w_S, \nabla_{\mathcal{T}} w_S)_{L^2(\Omega)} - k^2 (w_S, w_S)_{L^2(\Omega)} - 2 \operatorname{Re} \int_{\Xi^I_{\mathcal{T}}} [\![ w_S ]\!]_N \cdot \{ \overline{\nabla_{\mathcal{T}}} w_S \} \, dS \\ - 2 \operatorname{Re} \int_{\Xi^{\mathcal{B}}_{\mathcal{T}}} \delta w_S \, \overline{\nabla_{\mathcal{T}}} w_S \cdot \mathbf{n} \, dS = 0. \end{aligned} \tag{4.17}$$

From (4.16) it follows

**(1)** $[\![\nabla_{\mathcal{T}} w_S]\!]_N = 0 \quad \text{on} \quad \Xi_{\mathcal{T}}^{\mathcal{I}}$,

**(2)** $\nabla_{\mathcal{T}} w_S \cdot \mathbf{n} = 0 \quad \text{on} \quad \Xi_{\mathcal{T}}^{\mathcal{B}}$,

**(3)** $[\![w_S]\!]_N = 0 \quad \text{on} \quad \Xi_{\mathcal{T}}^{\mathcal{I}}$,

**(4)** $w_S = 0 \quad \text{on} \quad \Xi_{\mathcal{T}}^{\mathcal{B}}$

and this implies $w_S \in H_0^2(\Omega) \cap S$ (in particular that $\nabla_{\mathcal{T}} w_S = \nabla w_S$ holds).

Combining this result with the equation (4.17) we get

$$\|\nabla w_S\|_{L^2(\Omega)}^2 - k^2 \|w_S\|_{L^2(\Omega)}^2 = 0. \tag{4.18}$$

Define $v_S^*(x) = \langle x, \nabla w_S \rangle := x^T \cdot \nabla w_S$. From the real part of equation (4.15) it follows

$$\begin{aligned} 0 &= \mathrm{Re}\Big(a_{\mathcal{T}}(w_S, v_S^*) - k^2 (w_S, v_S^*)_{L^2(\Omega)}\Big) + \mathrm{Re}\Big(a_{\mathcal{T}}(w_S, v_S - v_S^*) - k^2 (w_S, v_S - v_S^*)_{L^2(\Omega)}\Big) \\ &\geq \mathrm{Re}\Big(a_{\mathcal{T}}(w_S, v_S^*) - k^2 (w_S, v_S^*)_{L^2(\Omega)}\Big) - |a_{\mathcal{T}}(w_S, v_S^* - v_S)| - |k^2 (w_S, v_S^* - v_S)_{L^2(\Omega)}|. \end{aligned} \tag{4.19}$$

We start with the first term, apply the integration by parts in equation (3.18), and use so called DG magic formula:

$$\sum_{K \in \mathcal{T}} \int_{\partial K} v\sigma \cdot \mathbf{n} = \int_{\mathcal{F}_{\mathcal{T}}^{\mathcal{I}}} [\![v]\!]_N \cdot \{\sigma\} + \int_{\mathcal{F}_{\mathcal{T}}^{\mathcal{I}}} \{v\} \cdot [\![\sigma]\!]_N + \int_{\mathcal{F}_{\mathcal{T}}^{\mathcal{B}}} [\![v]\!]_N \cdot \{\sigma\}.$$

We get

$$\begin{aligned} a_{\mathcal{T}}(w_S, v_S^*) - k^2 (w_S, v_S^*)_{L^2(\Omega)} = (\Delta_{\mathcal{T}} w_S, v_S^*)_{L^2(\Omega)} &- \int_{\Xi_{\mathcal{T}}^{\mathcal{I}}} [\![w_S]\!]_N \cdot \{\overline{\nabla_{\mathcal{T}} v_S^*}\} dS \\ &+ \int_{\Xi_{\mathcal{T}}^{\mathcal{I}}} [\![\nabla w_S]\!]_N \cdot \{\overline{v_S^*}\} dS - \int_{\Xi_{\mathcal{T}}^{\mathcal{B}}} \delta w_S \, \overline{\nabla_{\mathcal{T}} v_S^* \cdot \mathbf{n}} dS \\ &+ \int_{\Xi_{\mathcal{T}}^{\mathcal{B}}} (1-\delta) \nabla w_S \cdot \mathbf{n} \, \overline{v_S^*} \, dS - \frac{1}{ik} \int_{\Xi_{\mathcal{T}}^{\mathcal{I}}} \beta [\![\nabla w_S]\!]_N [\![\overline{\nabla_{\mathcal{T}} v_S^*}]\!]_N \\ &- \frac{1}{ik} \int_{\Xi_{\mathcal{T}}^{\mathcal{B}}} \delta \nabla w_S \cdot \mathbf{n} \overline{\nabla_{\mathcal{T}} v_S^* \cdot \mathbf{n}} dS + ik \int_{\Xi_{\mathcal{T}}^{\mathcal{I}}} \alpha [\![w_S]\!]_N \cdot [\![\overline{v_S^*}]\!]_N \\ &+ ik \int_{\Xi_{\mathcal{T}}^{\mathcal{B}}} (1-\delta) w_S \, \overline{v_S^*} \, dS - k^2 (w_S, v_S^*)_{L^2(\Omega)}. \end{aligned} \tag{4.20}$$

From conditions (1) – (4) we deduce that all boundary terms in (4.20) vanish. Hence (4.20) reduces to

$$\begin{aligned} a_{\mathcal{T}}(w_S, v_S^*) - k^2 (w_S, v_S^*)_{L^2(\Omega)} &= (\Delta_{\mathcal{T}} w_S, v_S^*)_{L^2(\Omega)} - k^2 (w_S, v_S^*)_{L^2(\Omega)} \\ &= \underbrace{(\nabla w_S, \nabla_{\mathcal{T}} v_S^*)_{L^2(\Omega)}}_{(i)} - k^2 \underbrace{(w_S, v_S^*)_{L^2(\Omega)}}_{(ii)}. \end{aligned}$$

We need to compute (i) and (ii). Using integration by parts and applying the fact that $2\,\mathrm{Re}(w_S\,\nabla\overline{w_S}) = \nabla(|w_S|^2)$, we get

(i)

$$\begin{aligned}\mathrm{Re}\int_\Omega w_S\overline{\langle x,\nabla w_S\rangle} &= \frac{1}{2}\int_\Omega \left\langle x,\nabla|w_S|^2\right\rangle\\ &= \frac{1}{2}\int_{\partial\Omega} x\cdot\mathbf{n}|w_S|^2 - \frac{1}{2}\int_\Omega(\nabla\cdot x)|w_S|^2\\ &= -\frac{d}{2}\|w_S\|^2_{L^2(\Omega)},\end{aligned} \tag{4.21}$$

(ii) Note that $w_S \in H^2_0(\Omega)$ so $\nabla\nabla^T w_S$ exists. Hence,

$$\begin{aligned}\mathrm{Re}\int_\Omega \nabla w_S\cdot\nabla_{\mathcal{T}}\overline{\langle x,\nabla w_S\rangle} &= \mathrm{Re}\int_\Omega \nabla w_S\cdot\nabla x\nabla\overline{w_S} + \mathrm{Re}\int_\Omega \nabla w_S\cdot(\nabla\nabla\overline{w_S}^T)x^T\\ &= \|\nabla w_S\|^2_{L^2(\Omega)} + \frac{1}{2}\int_\Omega \nabla|\nabla w_S|^2 x^T\\ &= (1-\frac{d}{2})\|\nabla w_S\|^2_{L^2(\Omega)}.\end{aligned} \tag{4.22}$$

By subtracting (4.21) and (4.22) we obtain

$$\mathrm{Re}\left(a_{\mathcal{T}}(w_S,v_S^*) - k^2(w_S,v_S^*)_{L^2(\Omega)}\right) = (1-\frac{d}{2})\|\nabla w_S\|^2_{L^2(\Omega)} + \frac{d}{2}k^2\|w_S\|^2_{L^2(\Omega)}.$$

Taking into account the continuity of $a_{\mathcal{T}}$ and applying the Cauchy-Schwarz inequality in the equation (4.19) lead to (see also [43], [23])

$$\begin{aligned}0 \geq&(2-d)\|\nabla w_S\|^2_{L^2(\Omega)} + dk^2\|w_S\|^2_{L^2(\Omega)} - 2C_c\|w_S\|_{DG}\|v_S^* - v_S\|_{DG^+}\\ &- 2k^2\|w_S\|_{L^2(\Omega)}\|v_S^* - v_S\|_{L^2(\Omega)}\\ \geq&(2-d)\|\nabla w_S\|^2_{L^2(\Omega)} + dk^2\|w_S\|^2_{L^2(\Omega)} - 2C_cC_S\|w_S\|_{DG}\|w_S\|_{L^2(\Omega)}\\ &- 2C_S k\|w_S\|^2_{L^2(\Omega)}.\end{aligned} \tag{4.23}$$

By using the definition of DG-norm and taking into account that $w_S \in H^2_0(\Omega)\cap S$ it follows $\|w_S\|_{DG} = \|w_S\|_{\mathcal{H}}$. For $d = 1$, we get

$$\begin{aligned}0 \geq& \|w_S\|^2_{\mathcal{H}} - 2C_cC_S\|w_S\|_{\mathcal{H}}\|w_S\|_{L^2(\Omega)} - 2C_S k\|w_S\|^2_{L^2(\Omega)}\\ \geq&\left(1 - \frac{2C_cC_S}{k} - \frac{2C_S}{k}\right)\|w_S\|^2_{\mathcal{H}}.\end{aligned} \tag{4.24}$$

If $C_S \leq k/2(1+C_c)$ then $w_S = 0 \quad \text{in} \quad \Omega$.

For $d = 2$ we derive from equation (4.23)

$$0 \geq 2k^2 \|w_S\|^2_{L^2(\Omega)} - 2C_c C_S \|w_S\|_{DG} \|w_S\|_{L^2(\Omega)} - 2C_S k \|w_S\|^2_{L^2(\Omega)}. \tag{4.25}$$

We add (4.18) to (4.25) to obtain

$$0 \geq \|\nabla w_S\|^2_{L^2(\Omega)} + k^2 \|w_S\|^2_{L^2(\Omega)} - 2C_c C_S \|w_S\|_{DG} \|w_S\|_{L^2(\Omega)} - 2C_S k \|w_S\|^2_{L^2(\Omega)}.$$

The same argument as in (4.24) finishes the proof for $d = 2$. For the $3d$ case from (4.23) it follows

$$0 \geq -\|\nabla w_S\|^2_{L^2(\Omega)} + 3k^2 \|w_S\|^2_{L^2(\Omega)} - 2C_c C_S \|w_S\|_{DG} \|w_S\|_{L^2(\Omega)} - 2C_S k \|w_S\|^2_{L^2(\Omega)}. \tag{4.26}$$

We add (4.18) to (4.26) to obtain

$$0 \geq 2k^2 \|w_S\|^2_{L^2(\Omega)} - 2C_c C_S \|w_S\|_{DG} \|w_S\|_{L^2(\Omega)} - 2C_S k \|w_S\|^2_{L^2(\Omega)}.$$

The same argument as in (4.25) finishes the proof for $d = 3$. □

**Remark 4.5.** Conditions (3.12) and (4.4), in general are not sufficient for discrete stability. If condition (4.12) is violated the discrete system possibly becomes singular. A simple example can be constructed by considering $\Omega := \text{conv}\{(0,0)^\intercal, (1,0)^\intercal, (0,1)^\intercal\}$ and the mesh $\mathcal{T}$ consists of the single element $\{\Omega\}$. A (one-dimensional) space $S$ which satisfies condition (3.12) is defined by the span of the squared cubic bubble function, $S = \text{span}\{(27\lambda_1\lambda_2\lambda_3)^2\}$, where $\lambda_1 = \xi_1, \lambda_2 = \xi_2, \lambda_3 = 1 - \xi_1 - \xi_2$ and $0 \leq \xi_1 \leq 1, 0 \leq \xi_2 \leq 1 - \xi_1$. In this case, equation (4.14) reduces to

$$(\nabla w_S, \nabla v_S)_{L^2(\Omega)} - k^2 (w_S, v_S)_{L^2(\Omega)} = 0 \qquad \forall v_S \in S. \tag{4.27}$$

As $S$ is a one-dimensional space we get the following $1 \times 1$ system $(A - k^2 B)w = 0$, where $A = \int_{\widehat{K}} \nabla b_1 \cdot \nabla b_1 = 5.1125$, $B = \int_{\widehat{K}} b_1^2 = 0.0843$ and $b_1 = (27\lambda_1\lambda_2\lambda_3)^2$. Obviously, the value of $k = \sqrt{\frac{A}{B}}$ is a critical wavenumber where the system matrix becomes singular. For general finite-dimensional spaces $S$, condition (4.12) can be interpreted as a condition on the scale resolution. However, (4.12) is always satisfied in the following important cases:

(a) Typically in the UWVF, the discrete space $S$ consists of systems of (discontinuous) plane waves. In that setting, condition (4.12) is trivially satisfied as then $S \cap H_0^2(\Omega) = \{0\}$.

(b) DG-methods based on classical piecewise polynomials on simplicial meshes (no curved element boundaries) satisfy (4.12) automatically as $\langle x, \nabla w_S \rangle \in S$ (see Theorem 6.4).

## 4.4 Convergence Analysis

In this section we will present the main theorems on the convergence of the DG-problem (3.18).

**Remark 4.6** (consistency)**.** Assume $\Omega \subset \mathbb{R}^d$, $f \in L^2(\Omega)$ and $g \in L^2(\partial\Omega)$ be such that the exact solution of the Helmholtz equation satisfies $u \in H^2$. We denote by $u_S$ the solution of the DG problem (3.17). Then

$$a_{\mathcal{T}}(u - u_S, v) = k^2(u - u_S, v)_{L^2(\Omega)} \qquad \forall v \in S. \tag{4.28}$$

*Proof.* It is enough to prove that $u$ satisfies the equation (3.17). From the $H^2$-regularity of $u$ it follows that

$$[\![u]\!]_N = 0, \quad [\![\nabla u]\!]_N = 0, \quad \{\nabla u\} = \nabla u \quad \text{on} \quad \mathfrak{S}_{\mathcal{T}}^{\mathcal{I}}.$$

We multiply both sides of equation (2.4a) by a test function $v \in S$, integrate elementwise and take the sum over all elements and finally apply integration by parts. We get

$$\sum_{K\in\mathcal{T}} \left( \int_{\partial K} (-\nabla u \cdot \mathbf{n})\overline{v} + \int_K \nabla u \cdot \nabla \overline{v} \right) - \int_\Omega k^2 u \overline{v} = \int_\Omega f \overline{v}. \tag{4.29}$$

Using the definition of the jumps on the inner edges, one get

$$\begin{aligned} -\sum_{K\in\mathcal{T}} \int_{\partial K} (\nabla u \cdot \mathbf{n})\overline{v} &= -\int_{\mathfrak{S}_{\mathcal{T}}^{\mathcal{B}}} \nabla u \cdot \mathbf{n}\overline{v} dS - \int_{\mathfrak{S}_{\mathcal{T}}^{\mathcal{I}}} \nabla u \cdot [\![\overline{v}]\!]_N dS \\ &= -\int_{\mathfrak{S}_{\mathcal{T}}^{\mathcal{B}}} \delta \nabla u \cdot \mathbf{n}\overline{v} dS - \int_{\mathfrak{S}_{\mathcal{T}}^{\mathcal{B}}} (1-\delta)\nabla u \cdot \mathbf{n}\overline{v} dS \\ &\quad - \int_{\mathfrak{S}_{\mathcal{T}}^{\mathcal{I}}} \nabla u \cdot [\![\overline{v}]\!]_N dS. \end{aligned}$$

We insert the boundary condition from equation (2.4b),

$$\begin{aligned}
-\sum_{K\in\mathcal{T}}\int_{\partial K}(\nabla u\cdot\mathbf{n})\overline{v} &= -\int_{\Xi_{\mathcal{T}}^B}\delta\nabla u\cdot\mathbf{n}\overline{v}dS-\int_{\Xi_{\mathcal{T}}^B}(1-\delta)g\overline{v}dS\\
&\quad+\int_{\Xi_{\mathcal{T}}^B}ik(1-\delta)u\overline{v}dS-\int_{\Xi_{\mathcal{T}}^I}\nabla u\cdot[\![\overline{v}]\!]_N dS\\
&= -\int_{\Xi_{\mathcal{T}}^B}\delta\nabla u\cdot\mathbf{n}\overline{v}dS-\int_{\Xi_{\mathcal{T}}^B}(1-\delta)g\overline{v}dS+\int_{\Xi_{\mathcal{T}}^B}ik(1-\delta)u\overline{v}dS\\
&\quad-\int_{\Xi_{\mathcal{T}}^I}\nabla u\cdot[\![\overline{v}]\!]_N dS+\frac{1}{ik}\int_{\Xi_{\mathcal{T}}^B}\delta g\overline{\nabla_{\mathcal{T}}v\cdot\mathbf{n}}dS\\
&\quad-\frac{1}{ik}\int_{\Xi_{\mathcal{T}}^B}\delta\nabla u\cdot\mathbf{n}\overline{\nabla_{\mathcal{T}}v\cdot\mathbf{n}}dS-\int_{\Xi_{\mathcal{T}}^B}\delta u\overline{\nabla_{\mathcal{T}}v\cdot\mathbf{n}}dS.
\end{aligned}$$

Inserting this result into (4.29) leads to

$$a_{\mathcal{T}}(u,v)-k^2(u,v)=(f,v)-\int_{\Xi_{\mathcal{T}}^B}\delta\frac{1}{ik}g\overline{\nabla_{\mathcal{T}}v\cdot\mathbf{n}}dS+\int_{\Xi_{\mathcal{T}}^B}(1-\delta)g\overline{v}dS\qquad\forall v\in S.$$

On the other hand, $u_S$ is the solution of the DG problem. The consistency (4.28) follows by subtracting the latter equation from (3.18). □

**Proposition 4.7.** *Let the exact solution of the weak problem (2.5) satisfy $u\in H^2(\Omega)$. Assume $0<\delta<1/3$, $C_S\le k/(2+2C_c)$ and*

$$\alpha_e>\frac{4d_{\mathcal{T}}}{3k}\max_{\iota\in\{+,-\}}C^2_{\mathrm{trace}}(S,K_e^\iota)\qquad\forall e\in\mathcal{F}_{\mathcal{T}}^{\mathcal{I},\mathrm{micro}}.$$

*where $d_{\mathcal{T}},C_S$ and $C_{\mathrm{trace}}$ were defined respectively in (3.5), (4.12) and (4.2). Denote by $u_S\in S$ the solution of the DG problem (3.17). Then*

$$\|u-u_S\|_{DG}\le\widetilde{C}\left(\inf_{v\in S}\|u-v\|_{DG^+}+\sup_{0\neq w_S\in S}\frac{k|(u-u_S,w_S)_{L^2(\Omega)}|}{\|w_S\|_{L^2(\Omega)}}\right),$$

*where $\widetilde{C}>0$ is a constant independent of $k$ and the mesh size.*

*Proof.* We start with a triangle inequality

$$\|u-u_S\|_{DG}\le\|u-v\|_{DG}+\|v-u_S\|_{DG}\qquad\forall v\in S\tag{4.30}$$

and employ the coercivity of $b_{\mathcal{T}}(\cdot,\cdot)$

$$\begin{aligned}
\|v-u_S\|^2_{DG}&\le\frac{1}{c_{\mathrm{coer}}}|b_{\mathcal{T}}(v-u_S,v-u_S)|\\
&\le\frac{1}{c_{\mathrm{coer}}}|b_{\mathcal{T}}(v-u,v-u_S)|+\frac{1}{c_{\mathrm{coer}}}|b_{\mathcal{T}}(u-u_S,v-u_S)|\\
&=\frac{1}{c_{\mathrm{coer}}}|b_{\mathcal{T}}(v-u,v-u_S)|+\frac{2k^2}{c_{\mathrm{coer}}}|(u-u_S,v-u_S)_{L^2(\Omega)}|,
\end{aligned}\tag{4.31}$$

where in the last inequality we used Remark 4.6.
The continuity of $b_{\mathcal{T}}(\cdot,\cdot)$ as in (4.7) together with (4.31) imply

$$\|v-u_S\|_{DG}^2 \leq \frac{C_c}{c_{\text{coer}}}\|v-u\|_{DG^+}\|v-u_S\|_{DG} + \frac{2k^2}{c_{\text{coer}}}|(u-u_S, v-u_S)_{L^2(\Omega)}|.$$

We combine this result with (4.30) and obtain

$$\begin{aligned}\|u-u_S\|_{DG} &\leq \|u-v\|_{DG} + \frac{C_c}{c_{\text{coer}}}\|v-u\|_{DG^+} + \frac{2k}{c_{\text{coer}}}\sup_{0\neq w_S\in S}\frac{|(u-u_S,w_S)_{L^2(\Omega)}|}{\|w_S\|_{L^2(\Omega)}}\\ &\leq (1+\frac{C_c}{c_{\text{coer}}})\|v-u\|_{DG^+} + \frac{2k}{c_{\text{coer}}}\sup_{0\neq w_S\in S}\frac{|(u-u_S,w_S)_{L^2(\Omega)}|}{\|w_S\|_{L^2(\Omega)}}. \qquad (4.32)\end{aligned}$$

□

**Proposition 4.8.** *([43, Proposition 8.1.4])*
*Let $\Omega\subset\mathbb{R}^2$ be a convex domain (or smooth and star-shaped). Then, the adjoint Helmholtz problem with right-hand side $w\in L^2(\Omega)$ :*
*Find $\phi\in H^1(\Omega)$ such that*

$$\int_\Omega \nabla\phi\nabla\bar{\psi}dV - k^2\int_\Omega \phi\bar{\psi}dV - ik\int_{\partial\Omega}\phi\bar{\psi}dS = \int_\Omega w\bar{\psi}dV \qquad \forall\psi\in H^1(\Omega) \qquad (4.33)$$

*has a unique solution. The corresponding solution operator is denoted by $N_k^*$, i.e., $\phi = N_k^* w$.*
*The solution $\phi$ belongs to $H^2(\Omega)$ and*

$$\|\phi\|_{\mathcal{H},\Omega} \leq C_1(\Omega)\|w\|_{L^2(\Omega)}, \qquad (4.34)$$
$$|\phi|_{2,\Omega} \leq C_2(\Omega)(1+k)\|w\|_{L^2(\Omega)}, \qquad (4.35)$$

*with $C_1(\Omega), C_2(\Omega) > 0$. By $|\phi|_{2,\Omega}$, we denote the $H^2(\Omega)$-seminorm containing only the second order derivatives.*

*Proof.* We recall the proof from [43] for the adjoint problem. Taking the test function $\psi = \phi$ in (4.33) and considering the real and imaginary part separately we get

$$|\phi|_{H^1(\Omega)}^2 - k^2\|\phi\|_{L^2(\Omega)}^2 \leq |\int_\Omega w\phi dV|, \qquad (4.36)$$
$$k\|\phi\|_{L^2(\partial\Omega)}^2 \leq |\int_\Omega w\phi dV|. \qquad (4.37)$$

From (4.37) and Young's inequality it follows

$$k\|\phi\|_{L^2(\partial\Omega)}^2 \leq \frac{1}{2\epsilon k}\|w\|_{L^2(\Omega)}^2 + \frac{\epsilon}{2}k\|\phi\|_{L^2(\Omega)}^2, \qquad (4.38)$$

Combining the latter inequality and (4.36) gives us

$$|\phi|^2_{H^1(\Omega)} \leq 2k^2\|\phi\|^2_{L^2(\Omega)} + \frac{1}{2}(1+k^{-2})\|w\|^2_{L^2(\Omega)}. \tag{4.39}$$

Now we choose the test function $\psi(x) = x^T \cdot \nabla\phi(x)$. From the equation (4.33) and Cauchy's inequality we get

$$k^2|\phi|^2_{L^2(\Omega)} \leq C(\Omega)\Big(k^2\|\phi\|^2_{L^2(\partial\Omega)} + \|w\|_{L^2(\Omega)}|\phi|_{H^1(\Omega)}\Big).$$

Using the results from (4.38) and (4.39) we get the following estimate

$$k^2|\phi|^2_{L^2(\Omega)} \leq C(\Omega)(1+k^{-2})\|w\|^2_{L^2(\Omega)}.$$

This result together with (4.39) gives the desired estimate in the $\mathcal{H}$-norm.
To get the estimate in the $H^2(\Omega)$-norm we use the regularity theory of $-\Delta$ (see [43]),

$$\begin{aligned}
|\phi|_{H^2(\Omega)} &\leq C(\Omega)\Big(|w + k^2\phi|_{L^2(\Omega)} + |\partial_n\phi|_{H^{1/2}(\partial\Omega)}\Big)\\
&\leq C(\Omega)\Big(|w|_{L^2(\Omega)} + k^2|\phi|_{L^2(\Omega)} + k|\phi|_{H^{1/2}(\Omega)}\Big)\\
&\leq C(\Omega)\Big(|w|_{L^2(\Omega)} + (1+k)\|\phi\|_{\mathcal{H}}\Big).
\end{aligned}$$

For the case of small $k$ we refer to [43]. □

**Proposition 4.9.** *Let the exact solution of the Helmholtz problem (2.5) satisfy $u \in H^2(\Omega)$. Assume $0 < \delta < 1/3$, $C_S \leq k/(2+2C_c)$ and*

$$\alpha_e > \frac{4d_{\mathcal{T}}}{3k}\max_{\iota\in\{+,-\}} C^2_{\text{trace}}(S, K^\iota_e) \qquad \forall e \in \mathcal{F}^{\mathcal{I},\text{micro}}_{\mathcal{T}},$$

*with $d_{\mathcal{T}}, C_S$ and $C_{\text{trace}}$ defined respectively in (3.5), (4.12) and (4.2). Then*

$$\sup_{0\neq w_S\in S}\frac{k|(u-u_S, w_S)_{L^2(\Omega)}|}{\|w_S\|_{L^2(\Omega)}} \leq C\eta_k(S)\Big(\inf_{v\in S}\|u-v\|_{DG^+} + \|u-u_S\|_{DG}\Big),$$

*where the adjoint approximation property is defined by*

$$\eta_k(S) := \sup_{f\in L^2(\Omega)\setminus\{0\}}\inf_{\psi_S\in S}\frac{k\|N^*_k f - \psi_S\|_{DG^+}}{\|f\|_{L^2(\Omega)}}. \tag{4.40}$$

*Proof.* The solution of the adjoint problem (4.33) with right-hand side $w_S \in S \subset L^2(\Omega)$ is denoted by $\phi$. The regularity estimates in Proposition 4.8 imply $\phi \in H^2(\Omega)$. From Remark 4.6 we get

$$(u-u_S, w_S)_{L^2(\Omega)} = a_{\mathcal{T}}(u-u_S, \phi) - k^2(u-u_S, \phi)_{L^2(\Omega)}.$$

Using the definition of the sesquilinear form $a_{\mathcal{T}}$ and the Galerkin orthogonality, we get for any $v \in S$

$$\begin{aligned}|(u-u_S, w_S)_{L^2(\Omega)}| &\le |a_{\mathcal{T}}(u-v, \phi-\psi_S)| + |a_{\mathcal{T}}(v-u_S, \phi-\psi_S)| \\ &\quad + k^2|(u-u_S, \phi-\psi_S)_{L^2(\Omega)}| \\ &\le (C_c\|u-v\|_{DG^+} + C_c\|v-u_S\|_{DG} \\ &\quad + \|u-u_S\|_{DG})\,\|\phi-\psi_S\|_{DG^+} \\ &\le (2C_c\|u-v\|_{DG^+} + (1+C_c)\|u-u_S\|_{DG})\,\|\phi-\psi_S\|_{DG^+}.\end{aligned}$$

From the latter inequality it follows

$$\frac{k|(u-u_S, w_S)_{L^2(\Omega)}|}{\|w_S\|_{L^2(\Omega)}} \le (2C_c\|u-v\|_{DG^+} + (1+C_c)\|u-u_S\|_{DG})\,\frac{k\|N_k^* w_S - \psi_S\|_{DG^+}}{\|w_S\|_{L^2(\Omega)}}.$$

Since $v, \psi_S$ are arbitrary functions in $S$ the following statement holds

$$\begin{aligned}\sup_{0\neq w_S\in S} \frac{k|(u-u_S, w_S)_{L^2(\Omega)}|}{\|w_S\|_{L^2(\Omega)}} &\le \Big(2C_c \inf_{v\in S}\|u-v\|_{DG^+} + (1+C_c)\|u-u_S\|_{DG}\Big) \\ &\quad \times \sup_{f\in L^2(\Omega)\setminus\{0\}} \inf_{\psi_S\in S} \frac{k\|N_k^* f - \psi_S\|_{DG^+}}{\|f\|_{L^2(\Omega)}}. \qquad (4.41)\end{aligned}$$

□

The combination of the previous results leads to the following wave-number explicit error estimate.

**Theorem 4.10.** *[quasi-optimal convergence]*
*Let the exact solution of the Helmholtz problem (2.5) satisfy $u \in H^2(\Omega)$. Assume $0 < \delta < 1/3$, $C_S \le k/(2+2C_c)$ and*

$$\alpha_e > \frac{4d_{\mathcal{T}}}{3k} \max_{\iota\in\{+,-\}} C^2_{\text{trace}}(S, K_e^\iota) \qquad \forall e \in \mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}},$$

*with $d_{\mathcal{T}}, C_S$ and $C_{\text{trace}}$ defined respectively in (3.5), (4.12) and (4.2). Then the condition*

$$\eta_k(S) < \frac{c_{\text{coer}}}{4(1+C_c)}$$

*implies the error estimate:*

$$\|u-u_S\|_{DG} \le C \inf_{v\in S}\|u-v\|_{DG^+},$$

*where $C$ is a constant independent of the choice of $k, h$ and the space $S$.*

*Proof.* By combining the results of Propositions 4.7 and 4.9, i.e. (4.32) and (4.41), we get the following:

$$\|u - u_S\|_{DG} \leq \left(1 + \frac{C_c}{c_{\text{coer}}} + \frac{4C_c}{c_{\text{coer}}}\eta_k(S)\right) \inf_{v \in S} \|u - v\|_{DG^+} + \frac{2(1+C_c)}{c_{\text{coer}}}\eta_k(S)\|u - u_S\|_{DG}.$$

The condition $2(1+C_c)\eta_k(S)/c_{\text{coer}} < 1/2$ allows us to absorb the second term in the right-hand side into the left-hand side. □

Later in Chapter 6 we estimate the adjoint approximation property $\eta_k(S)$ as well as $\inf_{v \in S} \|u - v\|_{DG^+}$ for the case of $hp$-FEM.

# 5

# Discrete Space

## 5.1 Piecewise Polynomials

We use the symbol $\nabla^n$ to denote derivatives of order $n$; more precisely, for a function $u : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^d$, we set

$$|\nabla^n u(x)|^2 = \sum_{\alpha \in \mathbb{N}_0^d : |\alpha| = n} \frac{n!}{\alpha!} |D^\alpha u(x)|^2.$$

The simplicial finite element mesh $\mathcal{T}$ consists of elements $K$ which are the images of the reference element $\widehat{K}$, i.e., the reference triangle (in 2D) or the reference tetrahedron (in 3D), under the element map $F_K : \widehat{K} \to K$.

**Assumption 5.1.** (shape-regular simplicial finite element mesh). Each element map $F_K$ can be written as $F_K = R_K \circ A_K$, where $A_K$ is an affine map (containing the scaling by $h_K$) and $R_K$ is analytic. Let $\widetilde{K} := A_K(K)$. The maps $R_K$ and $A_K$ satisfy for constants $C_{\text{affine}}, C_{\text{metric}}, \gamma > 0$ independent of $h$:

$$\begin{aligned} &\|A_K'\|_{L^\infty(\widehat{K})} \le C_{\text{affine}} h, \qquad &&\|(A_K')^{-1}\|_{L^\infty(\widehat{K})} \le C_{\text{affine}} h^{-1} \\ &\|(R_K')^{-1}\|_{L^\infty(\widetilde{K})} \le C_{\text{metric}}, \qquad &&\|\nabla^n R_K\|_{L^\infty(\widetilde{K})} \le C_{\text{metric}} \gamma^n n! \quad \forall n \in \mathbb{N}_0. \end{aligned}$$

**Remark 5.2.** Such kind of meshes can be obtained by a patchwise construction. Let $\mathcal{T}^{macro}$ be a fixed (coarse) mesh (with possibly curved elements) with analytic element maps that resolves the geometry. If the mesh $\mathcal{T}$ is obtained by quasi-uniform refinements of the reference element K and the final mesh is obtained by mapping the subdivisions of the reference element with the macro element maps, then the resulting element maps satisfy the Assumption 5.1.


**Figure 5.1**: Conforming and nonconforming meshes

**Definition 5.3.** *(i) A mesh is called conforming if the intersection of two elements is either empty, a vertex, a common edge or a common face (see Figure 5.1) and, additionally, any two elements $K, K' \in \mathcal{T}$ sharing a common face $e$ induce the same parametrization of $e$ via the element maps $F_K$ and $F_{K'}$.*

*(ii) A triangulation is called affine if every triangle can be transformed back to the reference triangle via an affine transformation.*

For meshes $\mathcal{T}$ satisfying Assumption 5.1 we define the conforming and nonconforming spaces of piecewise polynomials as follows:

$$S^{p,1}(\mathcal{T}) := \{u \in H^1(\Omega) | \quad \forall K \in \mathcal{T} : u|_K \circ F_K \in \mathcal{P}_p\}, \tag{5.1}$$

$$S^{p,0}(\mathcal{T}) := \{u \in L^2(\Omega) | \quad \forall K \in \mathcal{T} : u|_K \circ F_K \in \mathcal{P}_p\}, \tag{5.2}$$

where $\mathcal{P}_p^d$ denotes the space of polynomials of degree $p$, i.e,

$$\mathcal{P}_p^d = \text{span}\left\{ x_1^{i_1} \times x_2^{i_2} \times ... \times x_d^{i_d} : i_k \geq 0, \quad \sum_k i_k \leq p \right\}.$$

If $d$ is clear from the context we write $\mathcal{P}_p$ short for $\mathcal{P}_p^d$. Typically (5.1) requires conforming meshes. We recall some important inequalities for the $hp$-FEM which we need it later in the stability and convergence analysis. The following results are taken mainly from [59].

Let $D$ denotes a $d$-dimensional simplex with planar faces, so that there is an affine bijection which pulls $D$ the canonical $d$-dimensional simplex $T^d$, where

$$T^d = \left( (x_1, x_2, ..., x_d) \in \mathbb{R}^d : |x_i| \leq 1, \sum_{i=1}^{d} x_i \leq 2 - d \right).$$

Let $\{\psi_i\}_{i=1}^N$ be a $L^2(D)$-orthonormal basis, i.e.,

$$(\psi_i, \psi_j)_D = \int_D \psi_i(\mathbf{x})\psi_j(\mathbf{x})d\mathbf{x} = \delta_{ij}.$$

Then, any polynomial $u \in \mathcal{P}_p^d$ has a unique representation

$$u(\mathbf{x}) = \sum_{n=0}^{N-1} \widehat{u}_n \psi_n(\mathbf{x}),$$

where the coefficients are given by

$$\widehat{u}_n = \int_D u(\mathbf{x}) \psi_n(\mathbf{x}) d\mathbf{x}.$$

Orthogonal bases for the canonical $T^d$-simplex can be find in [21, 37, 55, 58] and have the form of products of Jacobi polynomials.

**Theorem 5.4.** *(Trace inequality for the planar triangle)*
*Let $K$ be a planar triangle. For any $v \in \mathcal{P}_p^2$ we have*

$$\|v\|_{L^2(\partial K)} \leq \sqrt{\frac{(p+1)(p+2)}{2} \frac{h_K}{|K|}} \|v\|_{L^2(K)}. \tag{5.3}$$

*For shape regular triangles it holds*

$$\frac{h_K}{|K|} \simeq h_K^{-1}.$$

*Proof.* First we consider the reference, right-angled, triangle,

$$T = (r, s | -1 \leq r, s \leq 1; r + s \leq 0).$$

An example for $L^2(T)$-orthonormal polynomial basis for $T$ with integer indices $i, j$ $(i \geq 0,\ j \geq 0, i + j \leq p)$ is

$$\psi_{i,j}(r,s) = P_i^{(0,0)}\left(\frac{2(r+1)}{1-s} - 1\right) \sqrt{\frac{2i+1}{2}} \left(\frac{1-s}{2}\right)^i P_j^{(2i+1,0)}(s) \sqrt{\frac{2(i+j)+2}{2}},$$

where $P_n^{(\alpha,\beta)}(x)$ is the Jacobi polynomial of order $n$ (see [59] and the references therein). We employ the expansion

$$v(r,s) = \sum_{i,j} \hat{v}_{ij} \psi_{i,j}(r,s).$$

For $s = -1$ it follows

$$\int_{-1}^{1} v^2(r,-1) dr = \hat{\mathbf{v}}^T F \hat{\mathbf{v}}.$$

Here the edge mass matrix has the following entries

$$F_{(ij),(kl)} = \int_{-1}^{1} \psi_{i,j}(r,-1)\psi_{k,l}(r,-1) dr \qquad = \delta_{ik}(-1)^{j+l} \sqrt{(i+j+1)(k+l+1)}.$$

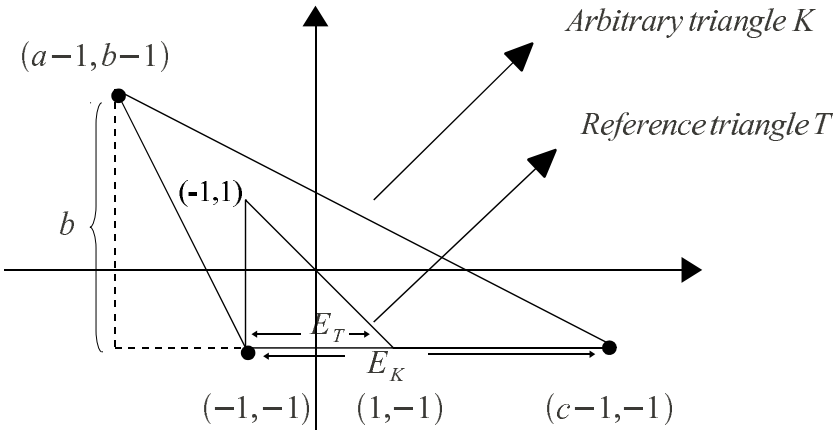


**Figure 5.2**: Reference and arbitrary triangles

Taking into account that $F$ is block diagonal, we can find the spectral radius of $F$

$$\rho(F) = \frac{1}{2}(p+1)(p+2),$$

so

$$\int_{-1}^{1} v^2(r,-1)dr \le \frac{1}{2}(p+1)(p+2)\|v\|_T^2. \tag{5.4}$$

Next we consider a general triangle $K$. Since integrals are invariant with respect to translations and rotations, we may choose the coordinate system in such a way that the vertices of $K$ have the coordinates $(-1,-1)^T, (c-1,-1)^T, (a-1,b-1)^T$, with some $a>0, b>0, c>0$ (see Figure 5.2). The edge $(-1,-1)^T(c-1,-1)^T$ is denoted by $E_K$.

$$\begin{pmatrix} r' \\ s' \end{pmatrix} = F\begin{pmatrix} r \\ s \end{pmatrix} = \begin{pmatrix} \frac{c}{2}+\frac{a}{2}-1 \\ \frac{b}{2}-1 \end{pmatrix} + \begin{pmatrix} \frac{c}{2} & \frac{a}{2} \\ 0 & \frac{b}{2} \end{pmatrix}\begin{pmatrix} r \\ s \end{pmatrix}.$$

Note that $\det F = cb/4$. For some $v \in \mathcal{P}_p^2$ we set $\tilde{v} = v \circ F$. Then

$$\int_K v^2 = \frac{cb}{4}\int_T \tilde{v}^2,$$
$$\int_{E_K} v^2 = \frac{c}{2}\int_{E_T} \tilde{v}^2.$$

By applying inequality (5.4) we derive

$$\int_{E_K} v^2 = \frac{c}{2}\int_{E_T} \tilde{v}^2 \le \frac{c}{2}\begin{pmatrix} p+2 \\ 2 \end{pmatrix}\int_T \tilde{v}^2 = \frac{2}{b}\begin{pmatrix} p+2 \\ 2 \end{pmatrix}\int_T \tilde{v}^2.$$

Finally, note that $1/b \leq ch_K/|K|$, where $c$ only depends on the shape regularity of the mesh.
The trace inequality (5.3) can be obtained by a rotation of the coordinate system in order to get the estimate for any edge. □

In general we have the following result from [59].

**Theorem 5.5.** *(Trace Inequality for the d-Simplex)*
*Let $D$ be a $d$-Simplex. For any $v \in \mathcal{P}_p^d$ we have*

$$\|v\|_{L^2(\partial D)} \leq \sqrt{\frac{(p+1)(p+d)}{d}\frac{Volume(\partial D)}{Volume(D)}}\|v\|_{L^2(D)}. \tag{5.5}$$

*Proof.* For a proof we refer to [59].

# 6

# Application to $hp$-Finite Elements

## 6.1 $hp$-FEM for Conforming Galerkin Discretization

In this section we present the estimates for the approximation property $\eta(S)$ and the convergence estimates for $hp$-finite elements of the problem (2.5):
Find $u \in H^1(\Omega)$ such that

$$\int_\Omega \left(\nabla u \nabla \bar{v} - k^2 u \bar{v}\right) dV + \mathrm{i}k \int_{\partial\Omega} u \bar{v} dS = \int_\Omega f \bar{v} dV + \int_{\partial\Omega} g \bar{v} dS \quad \forall v \in H^1(\Omega).$$

**Proposition 6.1.** *Let $\Omega \subset \mathbb{R}^d$, $d = 2,3$ be a bounded Lipschitz domain (with analytic boundary or a polygonal domain). Consider the problem (2.5) with $f \in L^2(\Omega)$ and $g \in H^{1/2}(\partial\Omega)$. The Galerkin discretization is defined by*

$$a(u_S, v) + b(u_S, v) = F(v) \qquad \forall v \in S := S^{p,1}(\mathcal{T}),$$

*where $a(\cdot,\cdot), b(\cdot,\cdot)$ and $F(\cdot)$ are as in (2.13)-(2.15). Let Assumption 5.1 hold and $k \geq k_0 > 1$. Then the following estimate holds*

$$\eta(S) \leq C \left\{ k^{3/2} \left( \left( \frac{h}{h+\sigma} \right)^p + k \left( \frac{kh}{\sigma p} \right)^p \right) + \frac{h}{p} \right\} \left( 1 + \frac{kh}{p} \right). \tag{6.1}$$

*Proof.* For a proof we refer to [47]. □

**Theorem 6.2.** *([47, Theorem 5.8, Corollary 5.10])*
*Let $\Omega \subset \mathbb{R}^d$, $d = 2,3$ be a bounded Lipschitz domain (with analytic boundary or a polygonal domain). Consider the model problem (2.4) with $f \in L^2(\Omega)$ and $g \in H^{1/2}(\partial\Omega)$. Let the Assumption 2.4 hold and $k \geq k_0 > 1$. Then there exist constants $c_1$ and $c_2$ independent of the mesh size $h$, polynomial degree $p$ and the wave number $k$ such that*

*(i)*

$$\frac{kh}{p} \le c_1 \quad \textit{together with} \quad p \ge 1 + c_2 \log(k) \tag{6.2}$$

*imply the following estimates*

$$\|u - u_S\|_{\mathcal{H},\Omega} \le 2(1 + C_b) \inf_{v \in S} \|u - v\|_{\mathcal{H},\Omega} \tag{6.3}$$

$$\|u - u_S\|_{L^2(\Omega)} \le C \frac{h}{p} \inf_{v \in S} \|u - v\|_{\mathcal{H},\Omega} \tag{6.4}$$

*where the constants $C_b$ and $C$ that are independent of $h, p, k$ and $f, g$.*

*(ii) Define $C_{f,g} := \|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}$. If $p \ge 1 + c_2 \log(k)$, then the condition $kh/p \le c_1$ implies the existence of the discrete solution and the a priori estimate*

$$\|u - u_S\|_{\mathcal{H},\Omega} \le C_{f,g} \frac{h}{p}. \tag{6.5}$$

*Proof.* For a proof we refer to Theorem 5.8 and Corollary 5.10 in [47]. □

## 6.2 *hp*–FEM for Discontinuous Galerkin Discretization

Theorem 4.10 provides a quasi-optimal error estimate for abstract approximation spaces $S$ which satisfy the condition (3.12). The concrete choice of the space $S$ enters the analysis via

(a) the constant $C_{\text{trace}}(S, K)$,

(b) the estimate of the approximation error $\inf_{v \in S} \|u - v\|_{DG^+}$,

(c) the adjoint approximation property $\eta_k(S)$.

In this section we derive explicit estimates for these quantities in the context of *hp*-finite element space which are explicit with respect to the polynomial degree $p$ and the mesh size $h$.

### 6.2.1 Discrete Stability

The estimate of $C_{\text{trace}}(S,K)$ in these cases is a local trace estimate for multivariate polynomials.

**Lemma 6.3.** *For the hp-finite element space* $S \in \left\{S^{p,1}, S^{p,0}\right\}$ *it holds*

$$C_{\text{trace}}(S,K) \leq \frac{cp}{\sqrt{h_K}}$$

*and the choice of* $\alpha|_e$ *as in (4.4) follows from condition (4.3).*

*Proof.* It is a direct consequence of Theorem 5.4. □

**Theorem 6.4.** *Let* $\alpha, \beta$ *and* $\delta$ *be chosen as follows*

$$\alpha_e = O\left(\max_{s\in\{+,-\}} \frac{p^2}{kh_{K_e^s}}\right), \quad \beta = O(\frac{kh}{p}), \quad \delta = \quad O(\frac{kh}{p}).$$

*Let* $S \in \{S^{p,1}, S^{p,0}\}$ *be the hp-finite element space corresponding to some mesh* $\mathcal{T}$ *which satisfies Assumption 5.1.*

- *If* $C_S$ *satisfies condition (4.12) then the UWVF has a unique solution in* $S$.
- *If* $\mathcal{T}$ *is an affine, shape regular triangulation of* $\Omega$ *then the UWVF has a unique solution in* $S$.

*Proof.* The first claim is simply the result of the Theorem 4.4. For the second part we show that $c_S \leq k/(1+C_c)$. First we recall the definition of the constant $C_S$

$$C_S = \sup_{w_S \in S\cap H_0^2(\Omega)\setminus\{0\}} \inf_{v_S\in S} \frac{\|\langle x, \nabla w_S\rangle - v_S\|_{DG^+}}{\|w_S\|_{L^2(\Omega)}}.$$

In this definition we can choose for any $w_S \in S \cap H_0^2(\Omega)\setminus\{0\}$ the test function $v_S := \langle x, \nabla w_S\rangle \in S$. From this choice it follows that $C_S = 0$ and by using the first part of this theorem we get the unique solvability of the method. □

### 6.2.2 Convergence Analysis

#### 6.2.2.1 General Non-Conforming $hp$-Finite Elements

In this section we consider general non-conforming $hp$-finite elements and we estimate the adjoint approximation property $\eta_k(S)$ and the error estimate for this case.

**Theorem 6.5.** *Let $\Omega \subset \mathbb{R}^d$, $d = 2, 3$ be a bounded Lipschitz domain with analytic boundary. Let $\alpha, \beta, \delta$ be chosen as follows*

$$\alpha_e = O\left(\max_{s\in\{+,-\}} \frac{p^2}{kh_{K_e^s}}\right), \quad \beta = O(\frac{kh}{p}), \quad \delta = \quad O(\frac{kh}{p}).$$

*There exist constants $C, \sigma > 0$ such that, for every $f \in L^2(\Omega)$ and $g \in H^{1/2}(\partial\Omega)$, it holds*

$$\inf_{w\in S} \|u - w\|_{DG^+} \le C_{f,g}\left\{k^\vartheta\left(1 + \frac{1}{\sqrt{p}}\left(\frac{kh}{p}\right)^{1/2} + \frac{kh}{p}\right)\left\{\frac{h}{p} + \left(\frac{kh}{\sigma p}\right)^p\right\} + \frac{h}{\sqrt{p}} + k\left(\frac{h}{p}\right)^2\right\},$$

*where $C_{f,g} := \|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}$.*

*Proof.* We employ the splitting of the solution of the Helmholtz equation $u = u_{H^2} + u_{\mathcal{A}}$ as in Theorem 2.5 with $u_{H^2} \in H^2(\Omega)$ and the analytic part $u_{\mathcal{A}}$. From the results of this theorem we can obtain the following estimates for the $H^2$ and analytic part of the solution,

$$\|u_{H^2}\|_{H^2(\Omega)} \le C\left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\right), \tag{6.6}$$

$$\|\nabla^p u_{\mathcal{A}}\|_{L^2(\Omega)} \le C\,(\lambda k)^{p-1} k^\vartheta \left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\right) \quad \forall p \in \mathbb{N}_0. \tag{6.7}$$

We approximate $u_{H^2}$ and $u_{\mathcal{A}}$ separately. From similar argument as in [46, Theorem 5.5], we conclude that there is $w_{H^2} \in S^{p,1}(\mathcal{T})$ such that for $q = 0, 1$, it holds

$$\|u_{H^2} - w_{H^2}\|_{H^q(K)} \le C\left(\frac{h_K}{p}\right)^{2-q} \|u_{H^2}\|_{H^2(K)} \qquad \forall K \in \mathcal{T}.$$

A summation over all elements leads to

$$k\|u_{H^2} - w_{H^2}\|_{\mathcal{H}} \le C\left(\frac{kh}{p} + \left(\frac{kh}{p}\right)^2\right)\left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\right). \tag{6.8}$$

From, e.g., [46, Lemma B.3], we know that for every $p$ there exists a bounded linear operator $\pi_p : H^s(\widehat{K}) \to \mathcal{P}_p$ with $s > d/2$, such that

$$\|u - \pi_p u\|_{H^t(\widehat{e})} \le C p^{-(s-1/2-t)} |u|_{H^s(\widehat{K})} \quad \text{for} \quad 0 \le t \le s - 1/2, \quad s > 1, \tag{6.9}$$

for each $t \in [0, s]$. Here, the constant $C > 0$ depends only on $t, s$. By $\widehat{K}$ we denote the reference element and by $\widehat{e}$ one of its edges (in 2D) resp. faces (in 3D).
By scaling this result to the mesh $\mathcal{T}$, we get the following estimates on the element $K$,

$$\|u - \pi_p u\|_{L^2(e)} \le C h_K^{3/2} p^{-3/2} |u|_{H^2(K)}, \tag{6.10}$$

$$\|u - \pi_p u\|_{H^1(e)} \le C h_K^{1/2} p^{-1/2} |u|_{H^2(K)}. \tag{6.11}$$

For the estimate of the boundary terms in the $DG^+$-norm we employ the definitions as in (4.4):

$$\alpha|_e = O(\max_{s\in\{+,-\}} \frac{p^2}{kh_{K_e^s}}) \quad (\text{with} \quad \alpha|_e > \frac{4}{3} d_{\mathcal{T}} \max_{s\in\{+,-\}} \frac{p^2}{kh_{K_e^s}}) \quad \forall e \in \mathcal{F}_{\mathcal{T}}^{\mathcal{I},\text{micro}},$$
$$\beta = O(\frac{kh}{p}), \quad \delta = \quad O(\frac{kh}{p}).$$

On the inner skeleton $\Im_{\mathcal{T}}^I$ we get

$$k\|\alpha^{-1/2}\{\nabla_{\mathcal{T}}(u_{H^2} - w_{H^2})\}\|^2_{L^2(\Im_{\mathcal{T}}^I)} \le \sum_{K\in\mathcal{T}} \sum_{e\in\mathcal{E}^I(K)} \sum_{e'\in \text{sons}(e)} k\alpha|_{e'}^{-1} \|\{\nabla_{\mathcal{T}}(u_{H^2} - w_{H^2})\}\|^2_{L^2(e')}.$$

Inserting the definition of $\alpha$ we get

$$\begin{aligned} k\|\alpha^{-1/2}\{\nabla_{\mathcal{T}}(u_{H^2} - w_{H^2})\}\|^2_{L^2(\Im_{\mathcal{T}}^I)} &\le \sum_{K\in\mathcal{T}} \sum_{e\in\mathcal{E}^I(K)} \sum_{e'\in \text{sons}(e)} \frac{3k}{4d_{\mathcal{T}}} \left(\min_{s\in\{+,-\}} \frac{kh_{K_e^s}}{p^2}\right) \\ &\qquad \times \|\nabla((u_{H^2} - w_{H^2})\big|_K)\|^2_{L^2(e')} \\ &\le \sum_{K\in\mathcal{T}} \frac{3k}{4d_{\mathcal{T}}} \frac{kh_K}{p^2} \sum_{e\in\mathcal{E}^I(K)} \|\nabla((u_{H^2} - w_{H^2})\big|_K)\|^2_{L^2(e)}. \end{aligned}$$

We use (6.11) to get

$$\begin{aligned} k\|\alpha^{-1/2}\{\nabla_{\mathcal{T}}(u_{H^2} - w_{H^2})\}\|^2_{L^2(\Im_{\mathcal{T}}^I)} &\le \sum_{K\in\mathcal{T}} \frac{3}{4d_{\mathcal{T}}} \left(\frac{k^2h_K^2}{p^3}\right) \sum_{e\in\mathcal{E}^I(K)} |u_{H^2}|^2_{H^2(K)} \\ &\le C\frac{3}{4}\frac{k^2h^2}{p^3}|u_{H^2}|^2_{H^2(K)}. \end{aligned}$$

By combining this result with (6.6) it follows

$$k^{1/2}\|\alpha^{-1/2}\{\nabla_{\mathcal{T}}(u_{H^2} - w_{H^2})\}\|_{L^2(\Im_{\mathcal{T}}^I)} \le C\left(\frac{kh}{p^{3/2}}\right)\left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\right). \tag{6.12}$$

The following estimates can be obtained by similar arguments,

$$k^{1/2}\|\beta^{1/2}[\![\nabla_{\mathcal{T}}(u_{H^2} - w_{H^2})]\!]_N\|_{L^2\left(\Im_{\mathcal{T}}^{\mathcal{I}}\right)} \le C\left(\frac{kh}{p}\right)\left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\right), \tag{6.13}$$

$$k^{1/2}\|\delta^{1/2}\nabla_{\mathcal{T}}(u_{H^2} - w_{H^2})\cdot\mathbf{n}\|_{L^2\left(\Im_{\mathcal{T}}^{\mathcal{B}}\right)} \le C\left(\frac{kh}{p}\right)\left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\right), \tag{6.14}$$

On the inner skeleton $\Im_{\mathcal{T}}^I$ we get in similar fashion

$$k^3\|\alpha^{1/2}[\![u_{H^2} - w_{H^2}]\!]_N\|^2_{L^2\left(\Im_{\mathcal{T}}^I\right)} \le \sum_{K\in\mathcal{T}} \sum_{e\in\mathcal{E}^I(K)} \sum_{e'\in \text{sons}(e)} k^3\alpha\|\,(u_{H^2} - w_{H^2})\big|_K\|^2_{L^2(e')}.$$

we insert the definition of $\alpha$ into the latter equation

$$\begin{aligned} k^3\|\alpha^{1/2}[\![u_{H^2}-w_{H^2}]\!]_N\|^2_{L^2(\Xi^I_{\mathcal{T}})} &\le \sum_{K\in\mathcal{T}}\sum_{e\in\mathcal{E}^I(K)}\sum_{e'\in\mathrm{sons}(e)}\frac{4d_{\mathcal{T}}}{3}k^3\left(\max_{s\in\{+,-\}}\frac{p^2}{kh_{K^s_e}}\right)\\ &\qquad\times\|\left(u_{H^2}-w_{H^2}\right)\big|_K\|^2_{L^2(e')}\\ &\le \sum_{K\in\mathcal{T}}\frac{4d_{\mathcal{T}}}{3}k^3\frac{p^2}{kh_K}\sum_{e\in\mathcal{E}^I(K)}\|\left(u_{H^2}-w_{H^2}\right)\big|_K\|^2_{L^2(e)}. \end{aligned}$$

Thus, using (6.10) it follows

$$\begin{aligned} k^3\|\alpha^{1/2}[\![u_{H^2}-w_{H^2}]\!]_N\|^2_{L^2(\Xi^I_{\mathcal{T}})} &\le \sum_{K\in\mathcal{T}}\frac{4d_{\mathcal{T}}}{3}\frac{k^2h_K^2}{p}\sum_{e\in\mathcal{E}^I(K)}|u_{H^2}|^2_{H^2(K)}\\ &\le C\frac{k^2h_K^2}{p}|u_{H^2}|^2_{H^2(K)}. \end{aligned}$$

Finally combining this result with (6.6) leads to

$$k^{3/2}\|\alpha^{1/2}[\![u_{H^2}-w_{H^2}]\!]_N\|_{L^2(\Xi^I_{\mathcal{T}})} \le C\left(\frac{kh}{\sqrt{p}}\right)\left(\|f\|_{L^2(\Omega)}+\|g\|_{H^{1/2}(\partial\Omega)}\right). \tag{6.15}$$

With the same argument we also can prove

$$k^{3/2}\|(1-\delta)^{1/2}(u_{H^2}-w_{H^2})\|_{L^2(\Xi^{\mathcal{B}}_{\mathcal{T}})} \le C\left(\frac{kh}{p}\right)^{3/2}\left(\|f\|_{L^2(\Omega)}+\|g\|_{H^{1/2}(\partial\Omega)}\right). \tag{6.16}$$

Thus, summing over the estimates (6.8) and (6.12)-(6.16) we get the following approximation property for the $H^2$-part:

$$k\|u_{H^2}-w_{H^2}\|_{DG^+} \le C\left(\frac{kh}{\sqrt{p}}+\left(\frac{kh}{p}\right)^2\right)\left(\|f\|_{L^2(\Omega)}+\|g\|_{H^{1/2}(\partial\Omega)}\right). \tag{6.17}$$

In [46, Theorem 5.5] the approximation $w_{\mathcal{A}}\in S^{p,1}(\mathcal{T})$ for $u_{\mathcal{A}}$ is constructed in an element-by-element fashion. For $K\in\mathcal{T}$, let the constant $C_K$ by defined by

$$C_K^2 := \sum_{p\in\mathbb{N}_0}\frac{\|\nabla^p u_{\mathcal{A}}\|^2_{L^2(K)}}{(2\lambda k)^{2p}}.$$

We have

$$\begin{aligned} &\|\nabla^p u_{\mathcal{A}}\|_{L^2(K)} \le (2\lambda k)^p C_K \quad \forall p\in\mathbb{N}_0,\\ &\sum_{K\in\mathcal{T}}C_K^2 \le \frac{4}{3}\left(\frac{C}{\lambda k}\right)^2 k^{2\vartheta}\left(\|f\|^2_{L^2(\Omega)}+\|g\|^2_{H^{1/2}(\partial\Omega)}\right). \end{aligned} \tag{6.18}$$

For $q \in \{0,1,2\}$ we get the following estimate (see [46, Theorem 5.5])

$$\|u_{\mathcal{A}} - w_{\mathcal{A}}\|_{H^q(K)} \leq C h_K^{-q} C_K \left\{ \left( \frac{h_K}{h_K + \sigma} \right)^{p+1} + \left( \frac{k h_K}{\sigma p} \right)^{p+1} \right\}. \tag{6.19}$$

By summing over all elements, it follows ([46])

$$k \|u_{\mathcal{A}} - w_{\mathcal{A}}\|_{\mathcal{H}} \leq C k^{\vartheta} \left( \frac{1}{p} + \frac{kh}{p} \right) \left\{ \frac{kh}{p} + k \left( \frac{kh}{\sigma p} \right)^{p} \right\} \left( \|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)} \right). \tag{6.20}$$

For the boundary terms we apply a multiplicative trace inequality [18]

$$\|v\|^2_{L^2(\partial K)} \leq C \left( \|v\|_{L^2(K)} |v|_{H^1(K)} + h_K^{-1} \|v\|^2_{L^2(K)} \right), \tag{6.21}$$

to obtain

$$k \|\alpha^{-1/2} \{\nabla_{\mathcal{T}} (u_{\mathcal{A}} - w_{\mathcal{A}})\}\|^2_{L^2(\Xi^I_{\mathcal{T}})} \leq \sum_{K \in \mathcal{T}} \sum_{e \in \mathcal{E}^I(K)} \sum_{e' \in \operatorname{sons}(e)} k \alpha^{-1} \|\nabla_{\mathcal{T}} ((u_{\mathcal{A}} - w_{\mathcal{A}})|_K)\|^2_{L^2(e')}.$$

From the definition of $\alpha$ it follows

$$\begin{aligned}
&k \|\alpha^{-1/2} \{\nabla_{\mathcal{T}} (u_{\mathcal{A}} - w_{\mathcal{A}})\}\|^2_{L^2(\Xi^I_{\mathcal{T}})} \\
&\leq \sum_{K \in \mathcal{T}} \sum_{e \in \mathcal{E}^I(K)} \sum_{e' \in \operatorname{sons}(e)} \frac{3k}{4 d_{\mathcal{T}}} \left( \min_{s \in \{+,-\}} \frac{k h_{K^s_{e'}}}{p^2} \right) \|\nabla ((u_{\mathcal{A}} - w_{\mathcal{A}})|_K)\|^2_{L^2(e')} \\
&\leq \sum_{K \in \mathcal{T}} \frac{3k}{4 d_{\mathcal{T}}} \left( \frac{k h_K}{p^2} \right) \sum_{e \in \mathcal{E}^I(K)} \|\nabla ((u_{\mathcal{A}} - w_{\mathcal{A}})|_K)\|^2_{L^2(e)} \\
&\leq \sum_{K \in \mathcal{T}} \frac{3}{4 d_{\mathcal{T}}} \left( \frac{k^2 h_K}{p^2} \right) \sum_{e \in \mathcal{E}^I(K)} \left( \|\nabla (u_{\mathcal{A}} - w_{\mathcal{A}})\|_{L^2(K)} |\nabla (u_{\mathcal{A}} - w_{\mathcal{A}})|_{H^1(K)} + h_K^{-1} \|\nabla (u_{\mathcal{A}} - w_{\mathcal{A}})\|^2_{L^2(K)} \right).
\end{aligned}$$

By using the estimates in equation (6.19) we get

$$\begin{aligned}
k \|\alpha^{-1/2} \{\nabla_{\mathcal{T}} (u_{\mathcal{A}} - w_{\mathcal{A}})\}\|^2_{L^2(\Xi^I_{\mathcal{T}})} &\leq \sum_{K \in \mathcal{T}} \frac{3C}{4} \frac{C_K^2 k^2}{p^2 h_K^2} \left\{ \left( \frac{h_K}{h_K + \sigma} \right)^{p+1} + \left( \frac{k h_K}{\sigma p} \right)^{p+1} \right\}^2 \\
&\leq \sum_{K \in \mathcal{T}} \frac{C C_K^2 k^2}{p^2} \left\{ h_K \left( \frac{h_K}{h_K + \sigma} \right)^{p-1} + \frac{k}{p} \left( \frac{k h_K}{\sigma p} \right)^{p} \right\}^2
\end{aligned}$$

Finally equation (6.18) gives us

$$\begin{aligned}
k^{1/2} \|\alpha^{-1/2} \{\nabla_{\mathcal{T}} (u_{\mathcal{A}} - w_{\mathcal{A}})\}\|_{L^2(\Xi^I_{\mathcal{T}})} \leq & C \frac{k^{\vartheta}}{p^3} \left\{ \frac{h}{p} + k \left( \frac{kh}{\sigma p} \right)^{p} \right\} \\
& \times \left( \|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)} \right).
\end{aligned} \tag{6.22}$$

By the similar arguments we obtain the following estimates

$$k^{1/2}\|\beta^{1/2}[\![\nabla_{\mathcal{T}}(u_{\mathcal{A}}-w_{\mathcal{A}})]\!]_N\|_{L^2(\Xi_{\mathcal{T}}^{\mathcal{I}})} \leq \quad C\frac{k^{\vartheta}}{p^{3/2}}\left\{\frac{h}{p}+k\left(\frac{kh}{\sigma p}\right)^p\right\} \times\Big(\|f\|_{L^2(\Omega)}+\|g\|_{H^{1/2}(\partial\Omega)}\Big), \tag{6.23}$$

$$k^{1/2}\|\delta^{1/2}\nabla_{\mathcal{T}}(u_{\mathcal{A}}-w_{\mathcal{A}})\cdot\mathbf{n}\|_{L^2(\Xi_{\mathcal{T}}^{\mathcal{B}})} \leq \quad C\frac{k^{\vartheta}}{p^{3/2}}\left\{\frac{h}{p}+k\left(\frac{kh}{\sigma p}\right)^p\right\} \times\Big(\|f\|_{L^2(\Omega)}+\|g\|_{H^{1/2}(\partial\Omega)}\Big). \tag{6.24}$$

Again, with a similar argument we derive

$$k^3\|\alpha^{1/2}[\![u_{\mathcal{A}}-w_{\mathcal{A}}]\!]_N\|^2_{L^2(\Xi_{\mathcal{T}}^{\mathcal{I}})} \leq \sum_{K\in\mathcal{T}}\sum_{e\in\mathcal{E}^{\mathcal{I}}(K)}\sum_{e'\in\mathrm{sons}(e)} k^3\alpha\|\,(u_{\mathcal{A}}-w_{\mathcal{A}})|_K\,\|^2_{L^2(e')}.$$

By using the definition of $\alpha$ we get

$$\begin{aligned}
&k^3\|\alpha^{1/2}[\![u_{\mathcal{A}}-w_{\mathcal{A}}]\!]_N\|^2_{L^2(\Xi_{\mathcal{T}}^{\mathcal{I}})}\\
&\leq \sum_{K\in\mathcal{T}}\sum_{e\in\mathcal{E}^{\mathcal{I}}(K)}\sum_{e'\in\mathrm{sons}(e)}\frac{4d_{\mathcal{T}}}{3}k^3\left(\max_{s\in\{+,-\}}\frac{p^2}{kh_{K_e^s}}\right)\|\,(u_{\mathcal{A}}-w_{\mathcal{A}})|_K\,\|^2_{L^2(e')}\\
&\leq \sum_{K\in\mathcal{T}}\frac{4d_{\mathcal{T}}}{3}k^3\left(\frac{p^2}{kh_K}\right)\sum_{e\in\mathcal{E}^{\mathcal{I}}(K)}\|\,(u_{\mathcal{A}}-w_{\mathcal{A}})|_K\,\|^2_{L^2(e)}\\
&\leq \sum_{K\in\mathcal{T}}\frac{4d_{\mathcal{T}}}{3}\left(\frac{k^2p^2}{h_K}\right)\sum_{e\in\mathcal{E}^{\mathcal{I}}(K)}\Big(\|u_{\mathcal{A}}-w_{\mathcal{A}}\|_{L^2(K)}|u_{\mathcal{A}}-w_{\mathcal{A}}|_{H^1(K)}+h_K^{-1}\|u_{\mathcal{A}}-w_{\mathcal{A}}\|^2_{L^2(K)}\Big).
\end{aligned}$$

Now we use the estimates in equation (6.19) to get

$$k^3\|\alpha^{1/2}[\![u_{\mathcal{A}}-w_{\mathcal{A}}]\!]_N\|^2_{L^2(\Xi_{\mathcal{T}}^{\mathcal{I}})} \leq \sum_{K\in\mathcal{T}} Cp^2k^2\left\{h_K\left(\frac{h_K}{h_K+\sigma}\right)^{p-1}+\frac{k}{p}\left(\frac{kh_K}{\sigma p}\right)^p\right\}^2 C_K^2.$$

Finally from equation (6.18) it follows

$$k^{3/2}\|\alpha^{1/2}[\![u_{\mathcal{A}}-w_{\mathcal{A}}]\!]_N\|_{L^2(\Xi_{\mathcal{T}}^{\mathcal{I}})} \leq \quad Ck^{\vartheta}\left\{\frac{h}{p}+k\left(\frac{kh}{\sigma p}\right)^p\right\}\Big(\|f\|_{L^2(\Omega)}+\|g\|_{H^{1/2}(\partial\Omega)}\Big). \tag{6.25}$$

With the same argument we can estimate the last boundary term as well,

$$k^{3/2}\|(1-\delta)^{1/2}(u_{\mathcal{A}}-w_{\mathcal{A}})\|_{L^2(\Xi_{\mathcal{T}}^{\mathcal{B}})} \leq \quad Ck^{\vartheta}\,\frac{(kh)^{1/2}}{p}\left\{\frac{h}{p}+k\left(\frac{kh}{\sigma p}\right)^p\right\}\Big(\|f\|_{L^2(\Omega)}+\|g\|_{H^{1/2}(\partial\Omega)}\Big). \tag{6.26}$$

The approximation property for the analytic part with respect to the $DG^+$ norm then follows from the estimates (6.20) and (6.22)-(6.26)

$$k\|u_{\mathcal{A}} - w_{\mathcal{A}}\|_{DG^+} \leq Ck^{\vartheta}\left(1 + \frac{1}{\sqrt{p}}\left(\frac{kh}{p}\right)^{1/2} + \frac{kh}{p}\right)\times \\ \left\{\frac{kh}{p} + k\left(\frac{kh}{\sigma p}\right)^{p}\right\}\left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\right). \tag{6.27}$$

The combination of the two estimates (6.17) and (6.27) leads to

$$\|u - w\|_{DG^+} \leq \left\{k^{\vartheta}\left(1 + \frac{1}{\sqrt{p}}\left(\frac{kh}{p}\right)^{1/2} + \frac{kh}{p}\right)\left\{\frac{h}{p} + \left(\frac{kh}{\sigma p}\right)^{p}\right\} + \frac{h}{\sqrt{p}} + k\left(\frac{h}{p}\right)^{2}\right\} \\ \times\left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\right),$$

which completes the proof. □

**Corollary 6.6.** *Let $\Omega \subset \mathbb{R}^d$, $d = 2,3$ be a bounded Lipschitz domain with analytic boundary. Let $\alpha, \beta, \delta$ be chosen as follows*

$$\alpha_e = O\left(\max_{s\in\{+,-\}} \frac{p^2}{kh_{K_e^s}}\right), \quad \beta = O(\frac{kh}{p}), \quad \delta = \quad O(\frac{kh}{p}).$$

*There exist constants $C, \sigma > 0$ such that, for every $f \in L^2(\Omega)$ and corresponding adjoint solution $v := N_k^* f$, it holds*

$$\eta_k(S) \leq C\left\{k^{\vartheta}\left(1 + \frac{1}{\sqrt{p}}\left(\frac{kh}{p}\right)^{1/2} + \frac{kh}{p}\right)\left\{\frac{h}{p} + \left(\frac{kh}{\sigma p}\right)^{p}\right\} + \frac{h}{\sqrt{p}} + k\left(\frac{h}{p}\right)^{2}\right\}.$$

*Proof.* Note that $u = N_k^* f = \overline{N_k \overline{f}}$ holds so that the regularity estimates as in Lemma 2.5 hold verbatim. Note that $g = 0$ in this case. □

Finally, the convergence estimate for $hp$-FEM can be stated in the following theorem.

**Theorem 6.7.** *[Convergence Estimate]*
*Let $\Omega \subset \mathbb{R}^d$, $d = 2,3$ be a bounded Lipschitz domain with analytic boundary. Let $\alpha$, $\beta, \delta$ be chosen as follows*

$$\alpha_e = O\left(\max_{s\in\{+,-\}} \frac{p^2}{kh_{K_e^s}}\right), \quad \beta = O(\frac{kh}{p}), \quad \delta = \quad O(\frac{kh}{p}).$$

*Moreover, let $0 < \delta < 1/3$. Then, there exist constants $c_1, c_2 > 0$ independent of $k, h$ and $p$ such that*

$$\frac{kh}{\sqrt{p}} \leq c_1 \quad \textit{together with} \quad p \geq c_2 \log(k)$$

*imply the following error estimates:*

$$\|u-u_S\|_{DG} \le C \inf_{v\in S} \|u-v\|_{DG^+}, \tag{6.28}$$

$$\|u-u_S\|_{DG} \le C\frac{h}{\sqrt{p}}\left(\|f\|_{L^2(\Omega)}+\|g\|_{H^{1/2}(\partial\Omega)}\right), \tag{6.29}$$

*where $C$ is independent of $k,h$ and $p$.*

*Proof.* By adjusting the constant $\tilde{c}$ in $\alpha|_e = \tilde{c}p^2/(h_{K_e}k)$ we get via Lemma 6.3 $\alpha|_e > 4C^2_{\text{trace}}(S,K_e)/k$. Hence, the only condition to check in Theorem 4.10 to get the convergence estimate is

$$\eta_k(S) < \frac{c_{\text{coer}}}{4(1+C_c)}.$$

From Corollary 6.6 we have

$$\eta_k(S) \le C\left\{k^{\vartheta}\left(1+\frac{1}{\sqrt{p}}\left(\frac{kh}{p}\right)^{1/2}+\frac{kh}{p}\right)\left\{\frac{h}{p}+\left(\frac{kh}{\sigma p}\right)^{p}\right\}+\frac{h}{\sqrt{p}}+k\left(\frac{h}{p}\right)^{2}\right\}.$$

Now to bound the right hand side with $c_{\text{coer}}/4(1+C_c)$, we simply need to adjust the constants $c_1$ and $c_2$ in the following conditions

$$\frac{kh}{\sqrt{p}} \le c_1 \quad \text{and} \quad p \ge c_2 \log(k).$$

From the Theorem 4.10 we get

$$\|u-u_S\|_{DG} \le C \inf_{v\in S} \|u-v\|_{DG^+}.$$

The combination of this result with Theorem 6.5 completes the proof. □

**Remark 6.8.** (i) For conforming and nonconforming affine polynomial finite element spaces the ultra weak variational formulation is unconditionally stable.

(ii) In order to get a quantitative error estimate explicitly in terms of $k,h$ and $p$, the following resolution condition is needed:

$$\frac{kh}{\sqrt{p}} \le c_1 \quad \text{and} \quad p \ge c_2 \log(k).$$

(iii) The convergence rate is reduced by half a power of $p$ compared to best approximation result.

(iv) For an important subclass of non-conforming spaces as well as for conforming spaces the Situation can be improved and the best approximation rate can be obtained (cf. Section 6.2.2.2).

**Corollary 6.9.** *With the same assumptions as in Theorem 6.7, there exist constants $c_1, c_2 > 0$ independent of $k, h$ and $p$ such that*

$$\frac{kh}{p} \le c_1 \quad \textit{together with} \quad p \ge c_2 \log(k)$$

*imply the following estimates:*

$$\begin{aligned}
&\|\nabla_{\mathcal{T}}(u-u_S)\|_{L^2(\Omega)} \le C\frac{h}{\sqrt{p}}\Big(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\Big),\\
&\|[\![\nabla_{\mathcal{T}}(u-u_S)]\!]_N\|_{L^2\left(\Xi_{\mathcal{T}}^I\right)} \le Ch^{1/2}\Big(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\Big),\\
&\|[\![u-u_S]\!]_N\|_{L^2\left(\Xi_{\mathcal{T}}^I\right)} \le C\left(\frac{h}{p}\right)^{3/2}\Big(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\Big),
\end{aligned}$$

*where $C$ is independent of $k, h$ and $p$.*

If we have higher regularity for $f, g$ then we get the the following convegence estimate:

**Theorem 6.10** (Convergence Estimate)**.** *Let $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$, be a bounded Lipschitz domain with analytic boundary. Fix $s \in \mathbb{N}_0$. Let $\alpha, \beta, \delta$ be chosen as follows*

$$\alpha_e = \mathsf{a}\max_{s\in\{+,-\}}\frac{p^2}{kh_{K_e^s}}, \quad \beta = \mathsf{b}\frac{kh}{p}, \quad \delta = \quad \mathsf{d}\frac{kh}{p},$$

*with $\mathsf{a}$ sufficiently large. Moreover, let $0 < \delta < \frac{1}{3}$. Then, there exist constants $c_1$, $c_2$, $C > 0$ independent of $k, h$ and $p$ such that under the assumptions*

$$\frac{kh}{\sqrt{p}} \le c_1 \quad \textit{together with} \quad p \ge c_2 \log(k) \quad \textit{as well as} \quad p \ge s+1 \tag{6.30}$$

*there holds for $f \in H^s(\Omega)$ and $g \in H^{s+1/2}(\partial\Omega)$ a priori estimate*

$$\|u-u_S\|_{DG} \le C\left[\sqrt{p}\left(\frac{h}{p}\right)^{s+1} + k^{\vartheta-1}\left\{\left(\frac{h}{h+\sigma}\right)^p + k\left(\frac{kh}{\sigma p}\right)^p\right\}\right]\Big(\|f\|_{H^s(\Omega)} + \|g\|_{H^{s+1/2}(\partial\Omega)}\Big).$$

*In particular, under the additional assumption that $\mathsf{b}$ and $\mathsf{d}$ satisfy $\mathsf{b}, \mathsf{d} \ge c_0 > 0$, there holds*

$$\|\nabla_{\mathcal{T}}(u-u_S)\|_{L^2(\Omega)} + \sqrt{\frac{h}{p}}\|[\![\nabla_{\mathcal{T}}(u-u_S)]\!]_N\|_{L^2\left(\Xi_{\mathcal{T}}^I\right)} + \frac{p}{\sqrt{h}}\|[\![u-u_S]\!]_N\|_{L^2\left(\Xi_{\mathcal{T}}^I\right)} \le C\|u-u_S\|_{DG}.$$

*Proof.* cf. [45]. □

#### 6.2.2.2 Convergence analysis for conforming and close to conforming $hp$-finite element spaces

The *a priori* estimate in Theorem 6.10 is optimal in $h$ (note that $f \in H^s(\Omega)$ with $g \in H^{s+1/2}(\partial\Omega)$ implies $u \in H^{s+2}(\Omega)$ by the assumed smoothness of $\partial\Omega$) but suboptimal in $p$ by half an order. This suboptimality is also present in the scale resolution condition (6.30). This is typical of $p$-explicit DG methods. While this suboptimality is sharp in the general case, [26], it can be removed in the present case by assuming that the approximation space contains an $H^1$-conforming subspace that is sufficiently rich. In order to be able to define an $H^1$-conforming space $S$, we have to require that the element maps $F_K$ be compatible across faces. For future reference, we formulate this as an assumption:

**Assumption 6.11** (conforming case)**.** A triangulation $\mathcal{T}$ satisfying Assumption 2.4 is said to be *regular* if the triangulation has no hanging nodes or edges and, additionally, any two elements $K, K' \in \mathcal{T}$ sharing a common face $e$ induce the same parametrization of $e$ via the element maps $F_K$ and $F_{K'}$. An approximation space $S$ is said to fall into the conforming case, if

$$S \supset S^{p,1}(\mathcal{T}) := \{u \in H^1(\Omega) \,|\, \forall K \in \mathcal{T} : u|_K \circ F_K \in \mathcal{P}_p\}.$$

We note that the classical space $S^{p,1}(\mathcal{T})$ has good (local) approximation properties. We also note that $S^{p,0}(\mathcal{T}) \supset S^{p,1}(\mathcal{T})$.

In a setting where the approximation space $S$ contains $S^{p,1}(\mathcal{T})$, a key observation is that certain jumps can be forced to be zero. We formulate this in the following remark:

**Remark 6.12.** In the conforming case the approximants $w_{H^{s+2}}$ and $w_{\mathcal{A}}$ in Theorem 6.5 can be chosen to be in $H^1(\Omega)$ (see, e.g., [46, Proof of Thm. 5.5]) so the following boundary terms vanish

$$k^{3/2}\|\alpha^{1/2}[\![u_{H^{s+2}} - w_{H^{s+2}}]\!]_N\|_{L^2(\Xi_{\mathcal{T}}^I)} = 0$$
$$k^{3/2}\|\alpha^{1/2}[\![u_{\mathcal{A}} - w_{\mathcal{A}}]\!]_N\|_{L^2(\Xi_{\mathcal{T}}^I)} = 0.$$

This lead to the following estimates for $p \geq s+1$

$$k\|u_{H^{s+2}} - w_{H^{s+2}}\|_{DG^+} \leq C\left(\frac{h}{p}\right)^s \left(\frac{kh}{p} + \left(\frac{kh}{p}\right)^2\right)\left(\|f\|_{H^s(\Omega)} + \|g\|_{H^{s+1/2}(\partial\Omega)}\right),$$
$$k\|u_{\mathcal{A}} - w_{\mathcal{A}}\|_{DG^+} \leq Ck^{\vartheta}\left(\left(\frac{h}{h+\sigma}\right)^p + k\left(\frac{kh}{\sigma p}\right)^p\right)\left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\partial\Omega)}\right).$$

■

We get the following error estimate for the conforming case:

**Theorem 6.13.** *Let $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$ be a bounded Lipschitz domain with an analytic boundary. Assume the conforming case of Assumption 6.11. Fix $s \in \mathbb{N}_0$ and $\overline{C} > 0$. Let $\alpha, \beta, \delta$ be chosen as follows*

$$\alpha_e = \mathsf{a} \max_{s\in\{+,-\}} \frac{p^2}{kh_{K_e^s}}, \quad \beta = \mathsf{b}\frac{kh}{p}, \quad \delta = \quad \mathsf{d}\frac{kh}{p}.$$

*Assume $p \geq s+1$.*

*(i) The adjoint consistency $\eta_k(S)$ satisfies with $\theta$ given by (2.25):*

$$\eta_k(S) \leq C\left(\frac{kh}{p} + k^{\vartheta}\left\{\left(\frac{h}{h+\sigma}\right)^p + k\left(\frac{kh}{\sigma p}\right)^p\right\}\right),$$

*(ii) The solution $u$ of (2.4a), (2.4b) satisfies with $\widetilde{C}_{f,g} := \|f\|_{H^s(\Omega)} + \|g\|_{H^{s+1/2}(\partial\Omega)}$*

$$\inf_{v\in S} k\|u-v\|_{DG^+} \leq \widetilde{C}_{f,g}\left(\left(\frac{h}{p}\right)^s \frac{kh}{p} + k^{\vartheta}\left\{\left(\frac{h}{h+\sigma}\right)^p + k\left(\frac{kh}{\sigma p}\right)^p\right\}\right).$$

*(iii) Let $u$ solve (2.4a), (2.4b). Let $0 < \delta < \frac{1}{3}$. Then there exist constant $c_1$, $c_2$, $\mathsf{a}_0 > 0$ independent of $h$, $p$, and $k$ such that for $\mathsf{a} \geq \mathsf{a}_0$ the conditions*

$$\frac{kh}{p} \leq c_1 \quad \textit{together with} \quad p \geq c_2 \log(k) \quad \textit{as well as} \quad p \geq s+1$$

*imply that the Galerkin approximation $u_S$ exists and satisfies the estimate*

$$\|u-u_S\|_{DG} \leq C\left[\left(\frac{h}{p}\right)^{s+1} + k^{\vartheta-1}\left\{\left(\frac{h}{h+\sigma}\right)^p + k\left(\frac{kh}{\sigma p}\right)^p\right\}\right]\left(\|f\|_{H^s(\Omega)} + \|g\|_{H^{s+1/2}(\partial\Omega)}\right).$$

*Proof.* The proof follows from the same arguments as in the non-conforming case. The key observation is that Remark 6.12 allows one to sharpen the estimates of Theorem 6.10. □

Inspection of the procedure leading to Theorem 6.13 shows that it is not essential that the mesh $\mathcal{T}$ be regular. Certain setting with hanging nodes are admissible. It suffices that the approximation space $S$ contain a subspace that is sufficiently rich. As an example, we mention a simple setting of meshes with hanging nodes that are obtained by refining a regular mesh in the sense of Assumption 6.11. To fix ideas, we introduce the notion of triangulations that are "close to regular" as follows:

**Definition 6.14** (triangulations close to regular)**.** *Let $\widehat{\mathcal{T}}_1,\dots,\widehat{\mathcal{T}}_L$ be a fixed selection of affine (not necessarily regular) triangulations of the reference element $\widehat{K}$. A triangulation $\mathcal{T}'$ is said to be* "close to regular" *if there is a regular triangulation $\mathcal{T}$ of $\Omega$ (in the sense of Assumption 6.11) with element maps $F_K : \widehat{K} \to K$ that induces the triangulation $\mathcal{T}'$ in the following way: For each element $K \in \mathcal{T}$, one can select a triangulation $\widehat{\mathcal{T}}_i$, $i \in \{1,\dots,L\}$ such that the elements $K' \in \mathcal{T}'$ with $K' \subset K$ are the images of the elements of $\widehat{\mathcal{T}}_i$ under $F_K$. Furthermore, the element maps corresponding to the elements $K' \in \mathcal{T}'$ with $K' \subset K$ are a composition of $F_K$ with an affine map that maps the corresponding subsimplex of $\widehat{\mathcal{T}}_i$ to the reference element $\widehat{K}$.*

**Corollary 6.15.** *Let $\mathcal{T}'$ be a triangulation that is "close to regular" in the sense of Definition 6.14. regular mesh in the sense of Assumption 6.11. Then the assertions of Theorem 6.13 are still valid for the space $S = S^{p,0}(\mathcal{T}')$.*

*Proof.* If $\mathcal{T}$ is the triangulation that engenders $\mathcal{T}'$ in the sense of Definition 6.14, then clearly $S^{p,0}(\mathcal{T}') \supset S^{p,1}(\mathcal{T})$. Hence, the corollary follows. □

**Remark 6.16.** For the case of bounded domains with non analytical boundaries such as polygonal or polyhedral domains we expect the similar results.

# 7

# Plane Waves

## 7.1 Preliminaries

Let $d = 2$. We define the space of linear combinations of $p \in \mathbb{N}$ plane waves of wavelength $2\pi/k$ , $k > 0$, in $\mathbb{R}^2$, denoted by $PW_k(\mathbb{R}^2)$, as follows:

$$PW_k(\mathbb{R}^2) = \{v \in C^\infty(\mathbb{R}^2) : v(\mathbf{x}) = \sum_{j=1}^{p} \alpha_j \exp(ik\mathbf{d}_j \cdot \mathbf{x}), \alpha_j \in \mathbb{C}\}, \tag{7.1}$$

where $D = \{\mathbf{d}_1, ..., \mathbf{d}_p\} \subset \mathbb{R}^2$ is a finite set of pairwise different vectors (different directions) with unit length. In can be seen that the set of $\{\exp(ik\mathbf{d}_j \cdot)\}_{j=1}^{p}$ is a basis of $PW_k(\mathbb{R}^2)$ for all $k > 0$ (cf. [28]).
We also define the space of local plane waves on an element $K \in \mathcal{T}$ as follows:

$$V_p(K) = \{v \in L^2(K) : v(\mathbf{x}) = \sum_{j=1}^{p} \alpha_j \exp(ik\mathbf{d}_j \cdot \mathbf{x}), \alpha_j \in \mathbb{C}\}, \tag{7.2}$$

then

$$V_{h,p}(\mathcal{T}) = \{v \in L^2(\Omega) : v|_K \in V_p(K) \forall K \in \mathcal{T}\}. \tag{7.3}$$

## 7.2 Ultra Weak Variational Formulation

The choice of plane waves as discretization space has become popular in the recent years. Plane waves satisfy locally the homogeneous differential equation as follows

from partial integration:

$$\sum_{K\in\mathcal{T}}(\nabla u,\nabla v)_{L^2(K)}-k^2(u,v)_{L^2(K)}=\sum_{K\in\mathcal{T}}\int_{\partial K}u\overline{\nabla v}\cdot.\mathbf{n}$$
$$=\int_{\mathcal{F}_{\mathcal{T}}^{\mathcal{I}}}[\![u]\!]_N\cdot\{\overline{\nabla_{\mathcal{T}}v}\}\,dS+\int_{\mathcal{F}_{\mathcal{T}}^{\mathcal{I}}}\{u\}[\![\overline{\nabla_{\mathcal{T}}v}]\!]_N\,dS$$
$$+\int_{\mathcal{F}_{\mathcal{T}}^{\mathcal{B}}}[\![u]\!]_N\{\overline{\nabla_{\mathcal{T}}v}\}\,dS. \tag{7.4}$$

As a consequence, all volume terms are vanishing in the bilinear form $a_{\mathcal{T}}(\cdot,\cdot)$ and the problem is formulated only on the mesh skeleton:
Find $u\in H^2$ such that $\mathcal{A}_h(u,v_p)=l_h(v_p)\quad\forall v_p\in V_{h,p}(\mathcal{T})$, where

$$\mathcal{A}_h(u,v):=\int_{\mathcal{F}_{\mathcal{T}}^{\mathcal{I}}}\{u\}[\![\overline{\nabla_{\mathcal{T}}v}]\!]_N\,dS-\frac{1}{ik}\int_{\mathcal{F}_{\mathcal{T}}^{\mathcal{I}}}\beta[\![\nabla_{\mathcal{T}}u]\!]_N[\![\overline{\nabla_{\mathcal{T}}v}]\!]_N\,dS-\int_{\mathcal{F}_{\mathcal{T}}^{\mathcal{I}}}\{\nabla_{\mathcal{T}}u\}\cdot[\![\overline{v}]\!]_N\,dS$$
$$+ik\int_{\mathcal{F}_{\mathcal{T}}^{\mathcal{I}}}\alpha[\![u]\!]_N\cdot[\![\overline{v}]\!]_N\,dS+\int_{\mathcal{F}_{\mathcal{T}}^{\mathcal{B}}}(1-\delta)u\overline{\nabla_{\mathcal{T}}v\cdot\mathbf{n}}\,dS-\frac{1}{ik}\int_{\mathcal{F}_{\mathcal{T}}^{\mathcal{B}}}\delta\nabla_{\mathcal{T}}u\cdot\mathbf{n}\overline{\nabla_{\mathcal{T}}v\cdot\mathbf{n}}\,dS$$
$$-\int_{\mathcal{F}_{\mathcal{T}}^{\mathcal{B}}}\delta\nabla_{\mathcal{T}}u\cdot\mathbf{n}\overline{v}\,dS+ik\int_{\mathcal{F}_{\mathcal{T}}^{\mathcal{B}}}(1-\delta)u\overline{v}\,dS, \tag{7.5}$$

and

$$l_h(v)=-\frac{1}{ik}\int_{\mathcal{F}_{\mathcal{T}}^{\mathcal{B}}}\delta g\overline{\nabla_{\mathcal{T}}v\cdot\mathbf{n}}\,dS+\int_{\mathcal{F}_{\mathcal{T}}^{\mathcal{B}}}(1-\delta)g\overline{v}\,dS. \tag{7.6}$$

## 7.3 Stability and Convergence Analysis

In this section we restrict ourselves to the case of affine triangulation for simplicial piecewise polynomial finite element space and compare it with the recent known results for the plane wave finite element space.

### 7.3.1 Norms

In the convergence analysis we need the following norms defined on the skeleton of $\mathcal{T}$,

$$\|v\|^2_{\mathcal{F}_{\mathcal{T}}}:=k^{-1}\|\beta^{1/2}[\![\nabla_{\mathcal{T}}v]\!]_N\|^2_{L^2\left(\mathcal{F}_{\mathcal{T}}^{\mathcal{I}}\right)}+k\|\alpha^{1/2}[\![v]\!]_N\|^2_{L^2\left(\mathcal{F}_{\mathcal{T}}^{\mathcal{I}}\right)}+k^{-1}\|\delta^{1/2}\nabla_{\mathcal{T}}v\cdot\mathbf{n}\|^2_{L^2\left(\mathcal{F}_{\mathcal{T}}^{\mathcal{B}}\right)}$$
$$+k\|(1-\delta)^{1/2}v\|^2_{L^2\left(\mathcal{F}_{\mathcal{T}}^{\mathcal{B}}\right)}, \tag{7.7}$$
$$\|v\|^2_{\mathcal{F}_{\mathcal{T}}^{+}}:=\|v\|^2_{\mathcal{F}_{\mathcal{T}}}+k\|\beta^{-1/2}\{v\}\|^2_{L^2\left(\mathcal{F}_{\mathcal{T}}^{\mathcal{I}}\right)}+k^{-1}\|\alpha^{-1/2}\{\nabla_{\mathcal{T}}v\}\|^2_{L^2\left(\mathcal{F}_{\mathcal{T}}^{\mathcal{I}}\right)}$$
$$+k\|\delta^{-1/2}v\|^2_{L^2\left(\mathcal{F}_{\mathcal{T}}^{\mathcal{B}}\right)}. \tag{7.8}$$

### 7.3.2 Stability

**Remark 7.1.** (i) The plane wave discontinuous Galerkin method (PWDG) is well-posed. This is a direct result from the fact that the imaginary part of the bilinear form $\mathcal{A}_h(.,.)$ is in fact a norm (cf. [32]).

(ii) Our method is unconditionally stable for affine mesh and simplicial polynomial finite element spaces as well as for plane waves.

### 7.3.3 Convergence analysis

The following proposition states that the plane wave discretization for homogeneous Helmholtz problem is quasi-optimal without any resolution condition when we consider skeleton norms.

**Proposition 7.2.** *([32, Proposition 3.6]) Let $\Omega$ be a convex domain, and assume that each element $K \in \mathcal{T}$ is a convex Lipschitz domain. Assume that the mesh $\mathcal{T}$ is a quasi-uniform mesh*[1]*. Choose the flux parameters $\alpha,\beta$ and $\delta$ to be real, strictly positive and independent of $p,h$ and $k$ and $0 < \delta \leq 1/2$. Let $u$ be the analytical solution of the homogeneous Helmholtz problem and $u_p$ be the PWDG solution. Then, there exists a constant $C > 0$ independent of $h, p$ and $k$ such that*

$$\|u - u_p\|_{\mathcal{F}_\mathcal{T}} \leq C \inf_{v_p \in V_{h,p}(\mathcal{T})} \|u - v_p\|_{\mathcal{F}_\mathcal{T}^+}$$

The $\|\cdot\|_{\mathcal{F}_\mathcal{T}}$-norm controls only the error on the skeleton. Estimates in the $L^2$-norm can be derived; however, negative powers of $h$ appear in this case.

**Proposition 7.3.** *[32, Corollary 3.8] Let the same assumptions as in Proposition 7.2 hold. Then, there exists a constant $C > 0$ independent of $h, p$ and $k$ such that*

$$\|u - u_p\|_{L^2(\Omega)} \leq C diam(\Omega)(k^{-1/2}h^{-1/2} + k^{1/2}h^{1/2})\|u - u_p\|_{\mathcal{F}_\mathcal{T}}.$$

To estimate the error in a stronger norm, $H^1$-norm, it is proposed in [32] to solve a Neumann boundary value problem for $-\Delta + k^2$ by means of $p$-degree Lagrangian finite elements, thus $\mathcal{P}u_p$ can be obtained by solving a second order elliptic boundary value problem, where $\mathcal{P}$ is $H^1(\mathcal{T})$-orthogonal projection onto the space of globally continuous piecewise polynomials of degree at most $p$ . In our method the estimates in stronger norms are directly derived.

[1] A mesh is called quasi-uniform if there exists a constant $\rho \in (0,1)$ such that for each element $K \in \mathcal{T}$, $h_K \geq \rho h_\mathcal{T}$.

**Theorem 7.4.** *[32, Theorem 3.15 , Proposition 3.19 ] Let $u \in H^{l+1}(\Omega)$ be the analytical solution of the homogeneous Helmholtz problem and $u_p$ be the PWDG solution. For $p$ sufficiently large, there exists a $C = C(kh) > 0$ independent of $p$ and $u$, but depending on the product $kh$, such that*

$$\|u-u_p\|_{\mathcal{F}_\mathcal{T}} \leq Ck^{-1/2}h^{l-1/2}\left(\frac{\log(p)}{p}\right)^{l-1/2}\|u\|_{l+1,k,\Omega}, \tag{7.9}$$

$$k\|u-u_p\|_{L^2(\Omega)} \leq Cdiam(\Omega)h^{l-1}\left(\frac{\log(p)}{p}\right)^{l-1/2}\|u\|_{l+1,k,\Omega}, \tag{7.10}$$

$$k\|\nabla(u-\mathcal{P}(u_p))\|_{L^2(\Omega)} \leq C(diam(\Omega)+k^{-1})h^{l-1}\left(\frac{\log(p)}{p}\right)^{l-1/2}\|u\|_{l+1,k,\Omega}, \tag{7.11}$$

*where*

$$\|v\|^2_{l+1,k,\Omega} := \sum_{j=0}^{l+1} k^{2(l+1-j)}|v|^2_{H^j(\Omega)}.$$

It is also possible to define the flux parameters depending on $k$, $h$ and $p$ as follows:

$$\alpha = \frac{\mathrm{a}}{kh}\frac{p}{\log(p)}, \quad \beta = bkh\frac{\log(p)}{p}, \quad \delta = dkh\frac{\log(p)}{p},$$

where one can gain half a power of $\log(p)/p$ in the best approximation estimate compared to the case of constant flux parameters in $\mathcal{F}_\mathcal{T}$-norm, but at the end one get the same order of convergence in energy norm [32].

For the $h$-version of plane wave approximations the following result exists.

**Theorem 7.5.** *[28, Theorem 4.10 ] Let $\Omega$ be a convex domain (or smooth and star-shaped). Let $u$ be the analytical solution of the Helmholtz problem (2.4) and $u_h$ be the PWDG solution of the problem 3.17. Assume $\alpha = \boldsymbol{a}/(kh), \beta = bkh$ and $\delta = dkh$ with $\boldsymbol{a}, d > 0$ and $b \geq 0$ and $\delta \in (0, 1/2)$. Then there exist $a_0, c_0$ and $C = C(\Omega, p) > 0$ independent of $h$ and $k$ such that if $a \geq a_0$ and $k^2h \leq c_0$, then the following error bound is true*

$$\|u-u_h\|_{DG} \leq Ch\Big(k\|f\|_{L^2(\Omega)} + [c_0(h+c_0)]^{1/2}\|f-P_kf\|_{L^2(\Omega)}\Big),$$

*where $P_k$ denotes the $L^2(\Omega)$-orthogonal projection onto $V_{h,p}(\mathcal{T})$.*

From this theorem it can be seen that only the first order convergence rate can be derived (independent of the number of the plane waves which were used in the local approximation spaces) which is because of the fact that plane waves are solutions of homogeneous Helmholtz problem, so one can not get better than linear convergence (with respect to $h$) for an inhomogeneous problem. Also the condition $k^2h \leq c_0$ is very restrictive but in our method for an inhomogeneous Helmholtz problem when data are

regular enough, we derived the optimal convergence (6.28) as an $h$-version method (with a fixed $p$) under a milder condition.

**Remark 7.6.**

1. For the PWDG method one obtains quasi-optimality for the skeleton norm without any resolution condition for homogeneous Helmholtz problem as a $p$-version method.

2. The PWDG method also can be applied to inhomogeneous problems, while then the convergence rate is at most linear with respect to $h$. In addition one needs the restrictive resolution condition $k^2h < c_0$. Our method leads to an optimal rate of convergence (with respect to the DG norm) also for inhomogeneous problems under the conditions $kh/p \leq c_1$ and $p \geq c_2 \log(k)$.

3. To obtain error estimates for the PWDG method with respect to stronger norms, e.g. the $H^1$-norm, one has to solve an additional second order elliptic PDE, while this can be avoided for polynomial spaces.

4. The linear system in both methods are very ill-conditioned and indefinite. Typically, fast direct solvers are employed for its solution while iterative solvers for such highly indefinite problems are still in its infancies.

5. The algorithmic realization, the choice of efficient data structure and the quadrature problem for polynomial DG methods is well understood so that the implementation of the UWVF with polynomial finite elements can be performed in the framework of "standard polynomial finite element technology". The efficient realization of the PWDG is less standard, e.g., the quadrature for generating the system matrix is far from trivial and we refer to [15, 27, 28, 32] for the details.

# 8

# Conclusion

In this work we presented a DG method which is unconditionally stable for piecewise polynomials and plane waves. The unique solvability of the method for very general approximation spaces $S$ was also proved under a very mild resolution condition. We proved that the quasi optimality of the method for these abstract discretization spaces can be obtained under a certain condition on the approximation properties of the approximation space. The theory depends on the decomposition lemma [46, 47] which implies that the solution of the Helmholtz problem can be split into an $H^2-$part with „good" regularity constant and an oscillatory part which is analytic.

As an application of the general theory, we derived an error analysis for conforming and non-conforming $hp$- finite element spaces. We estimated the adjoint approximation property $\eta_k(S)$ explicitly in terms of the wave number $k$, mesh width $h$ and polynomial degree $p$, and got optimal convergence rates under some mild conditions on $k$, $h$ and $p$. We have considered a setting, where $f \in L^2(\Omega)$ so that the maximal regularity of the solution in general is $H^2$. For $hp$-finite elements it is well-known that a convergence rate of $O(h/p)$ with respect to the $H^1$-norm then is optimal. For more regular settings , we expect that our theory can be generalized and higher order convergence estimates can be proved.

# References


[1] J. D. Achenbach. *Wave propagation in elastic solids*. North Holland, 1973.

[2] M. Ainsworth, P. Monk, and W. Muniz. Dispersive and dissipative properties of discontinuous Galerkin finite element methods for the second-order wave equation. *J. Sci. Comput.*, 27(1-3):5–40, 2006.

[3] B. A. Auld. *Acoustic fields and waves in solids*. John Wiley and Sons, 1973.

[4] A. K. Aziz and R. B. Kellogg. A scattering problem for the Helmholtz equation. In *Advances in computer methods for partial differential equations, III (Proc. Third IMACS Internat. Sympos., Lehigh Univ., Bethlehem, Pa., 1979)*, pages 93–95. IMACS, New Brunswick, N.J., 1979.

[5] A. K. Aziz, R. B. Kellogg, and A. B. Stephens. A two point boundary value problem with a rapidly oscillating solution. *Numer. Math.*, 53(1-2):107–121, 1988.

[6] I. M. Babuška, U. Banerjee, and J.E. Osborn. Generalized finite element methods : Main ideas, results, and perspective. *Security*, 1(1):67–103, 2004.

[7] I. M. Babuška and S. A. Sauter. Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers? *SIAM J. Numer. Anal.*, 34(6):2392–2423, 1997.

[8] A. Buffa and P. Monk. Error estimates for the ultra weak variational formulation of the Helmholtz equation. *M2AN Math. Model. Numer. Anal.*, 42(6):925–940, 2008.

[9] O. Cessenat and B. Després. Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz problem. *SIAM J. Numer. Anal.*, 35(1):255–299 (electronic), 1998.

[10] O. Cessenat and B. Després. Using plane waves as base functions for solving time harmonic equations with the ultra weak variational formulation. *J. Comput. Acoust.*, 11(2):227–238, 2003. Medium-frequency acoustics.

[11] C. L. Chang. A least-squares finite element method for the Helmholtz equation. *Comput. Methods Appl. Mech. Engrg.*, 83(1):1–7, 1990.

[12] D. Colton and R. Kress. *Inverse acoustic and electromagnetic scattering theory*, volume 93 of *Applied Mathematical Sciences*. Springer-Verlag, Berlin, 1992.

[13] A. Craggs. The use of simple three-dimensional acoustic finite elements for determining the natural modes and frequencies of complex shaped enclosures. *J. Sound Vib.*, 23(3):331–339, 1972.

[14] P. Cummings and X. Feng. Sharp regularity coefficient estimates for complex-valued acoustic and elastic Helmholtz equations. *Math. Models Methods Appl. Sci.*, 16(1):139–160, 2006.

[15] L. Demkowicz, J. Gopalakrishnan, I. Muga, and J. Zitelli. Wavenumber explicit analysis of a DPG method for the multidimensional helmholtz equation. Technical Report 11-17, ICES, 2011.

[16] B. Després. Sur une formulation variationnelle de type ultra-faible. *C. R. Acad. Sci. Paris Sér. I Math.*, 318(10):939–944, 1994.

[17] Z. Ding. A proof of the trace theorem of Sobolev spaces on Lipschitz domains. *Proc. Amer. Math. Soc.*, 124(2):591–600, 1996.

[18] V. Dolejší, M. Feistauer, and C. Schwab. A finite volume discontinuous Galerkin scheme for nonlinear convection-diffusion problems. *Calcolo*, 39(1):1–40, 2002.

[19] Jim Douglas, Jr., Juan E. Santos, D. Sheen, and L. S. Bennethum. Frequency domain treatment of one-dimensional scalar waves. *Math. Models Methods Appl. Sci.*, 3(2):171–194, 1993.

[20] Jim Douglas, Jr., D. Sheen, and J. E. Santos. Approximation of scalar waves in the space-frequency domain. *Math. Models Methods Appl. Sci.*, 4(4):509–531, 1994.

[21] M. Dubiner. Spectral methods on triangles and other domains. *J. Sci. Comput.*, 6(4):345–390, 1991.

[22] S. Esterhazy and J.M. Melenk. On stability of discretizations of the Helmholtz equation. In T.Y. Lakkis O. Graham, I.G. Hou and R. Scheichl, editors, *Lecture Notes in Computational Science and Engineering*. Springer Verlag, 2011.

[23] X. Feng and H. Wu. Discontinuous Galerkin methods for the Helmholtz equation with large wave number. *SIAM J. Numer. Anal.*, 47(4):2872–2896, 2009.

[24] X. Feng and H. Wu. *hp*-discontinuous Galerkin methods for the Helmholtz equation with large wave number. *Math. Comp.*, 80(276):1997–2024, 2011.

[25] P. Filippi, D. Habault, J.P. Lefebvre, and A. Bergassoli. *Acoustics*. Academic Press, 1999.

[26] E. H. Georgoulis, E. Hall, and J. M. Melenk. On the suboptimality of the *p*-version interior penalty discontinuous Galerkin method. *J. Sci. Comput.*, 42(1):54–67, 2010.

[27] C. J. Gittelson. *Plane wave discontinuous Galerkin methods*. 2008. Thesis (M.Sc.)–ETH.

[28] C. J. Gittelson, R. Hiptmair, and I. Perugia. Plane wave discontinuous Galerkin methods: analysis of the *h*-version. *M2AN Math. Model. Numer. Anal.*, 43(2):297–331, 2009.

[29] P. Grisvard. *Elliptic problems in nonsmooth domains*, volume 24 of *Monographs and Studies in Mathematics*. Pitman (Advanced Publishing Program), Boston, MA, 1985.

[30] I. Harari. A survey of finite element methods for time-harmonic acoustics. *Comput. Methods Appl. Mech. Engrg.*, 195(13-16):1594–1607, 2006.

[31] I. Harari and T. J. R. Hughes. Galerkin/least-squares finite element methods for the reduced wave equation with nonreflecting boundary conditions in unbounded domains. *Comput. Methods Appl. Mech. Engrg.*, 98(3):411–454, 1992.

[32] R. Hiptmair, A. Moiola, and I. Perugia. Plane wave discontinuous Galerkin methods for the 2D Helmholtz equation: analysis of the *p*-version. *SIAM J. Numer. Anal.*, 49(1):264–284, 2011.

[33] J. T. Hunt, M. R. Knittel, and D. Barach. Finite-element approach to acoustic radiation from elastic structures. *J. Acoust. Soc. Am.*, 55(2):269–280, 1974.

[34] J. T. Hunt, M. R. Knittel, C. S. Nichols, and D. Barach. Finite-element approach to acoustic scattering from elastic structures. *J. Acoust. Soc. Am.*, 57(2):287–299, 1975.

[35] T. Huttunen and P. Monk. The use of plane waves to approximate wave propagation in anisotropic media. *J. Comput. Math.*, 25(3):350–367, 2007.

[36] F. Ihlenburg. *Finite element analysis of acoustic scattering*, volume 132 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1998.

[37] T. Koornwinder. Two-variable analogues of the classical orthogonal polynomials. In *Theory and application of special functions (Proc. Advanced Sem., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1975)*, pages 435–495. Math. Res. Center, Univ. Wisconsin, Publ. No. 35. Academic Press, New York, 1975.

[38] O. Laghrouche and P. Bettess. Solving short wave problems using special finite elements—towards an adaptive approach. In *The mathematics of finite elements and applications, X, MAFELAP 1999 (Uxbridge)*, pages 181–194. Elsevier, Oxford, 2000.

[39] O. Laghrouche, P. Bettess, and R. J. Astley. Modelling of short wave diffraction problems using approximating systems of plane waves. *International journal for numerical methods in engineering.*, 54(10):1501–1533, August 2002.

[40] L.D. Landau and E.M. Lifshitz. *Theory of elasticity*. Butterworth Heinemann, 1986.

[41] R. Leis. *Initial-boundary value problems in mathematical physics*. B. G. Teubner, Stuttgart, 1986.

[42] T. Luostari, T. Huttunen, and P. Monk. Plane wave methods for approximating the time harmonic wave equation. In *Highly oscillatory problems*, volume 366 of *London Math. Soc. Lecture Note Ser.*, pages 127–153. Cambridge Univ. Press, Cambridge, 2009.

[43] J. M. Melenk. *On generalized finite-element methods*. ProQuest LLC, Ann Arbor, MI, 1995. Thesis (Ph.D.)–University of Maryland, College Park.

[44] J. M. Melenk and I. Babuška. The partition of unity finite element method: basic theory and applications. *Comput. Methods Appl. Mech. Engrg.*, 139(1-4):289–314, 1996.

[45] J. M. Melenk, A. Parsania, and S. A. Sauter. Generalized DG methods for highly indefinite Helmholtz problems based on the ultra-weak variational formulation. Technical Report 03-2012, 2012.

[46] J. M. Melenk and S. A. Sauter. Convergence analysis for finite element discretizations of the Helmholtz equation with Dirichlet-to-Neumann boundary conditions. *Math. Comp.*, 79(272):1871–1914, 2010.

[47] J. M. Melenk and S. A. Sauter. Wave-number explicit convergence analysis for galerkin discretizations of the helmholtz equation. *SIAM J. Numer. Anal.*, 49:1210–1243, 2011.

[48] P. Monk, J. Schöberl, and A. Sinwel. Hybridizing Raviart-Thomas elements for the Helmholtz equation, 2010.

[49] P.M. Morse and K.U. Ingard. *Theoretical Acoustics*. McGraw-Hill Book Company, 1968.

[50] J.C. Nédélec. *Acoustic and electro magnetic equations*. Springer, 2000.

[51] P. Ortiz. Finite elements using a plane-wave basis for scattering of surface water waves. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 362(1816):525–540, 2004.

[52] E. Perrey-Debain, O. Laghrouche, P. Bettess, and J. Trevelyan. Plane-wave basis finite elements and boundary elements for three-dimensional wave scattering. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 362(1816):561–577, 2004.

[53] M. Petyt, J. Lea, and G. H. Koopmann. A finite element method for determining the acoustic modes of irregular shaped cavities. *J. Sound Vib.*, 45(4):495–502, 1976.

[54] A.D. Pierce. *Acoustics: An Introduction to its physical principles and applications*. The Acoustical Society of America, 1981.

[55] J. Proriol. Sur une famille de polynomes à deux variables orthogonaux dans un triangle. *C. R. Acad. Sci. Paris*, 245:2459–2461, 1957.

[56] M. Renardy and R. C. Rogers. *An introduction to partial differential equations*, volume 13 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1993.

[57] T. Strouboulis, I. Babuška, and R. Hidajat. The generalized finite element method for Helmholtz equation: theory, computation, and open problems. *Comput. Methods Appl. Mech. Engrg.*, 195(37-40):4711–4731, 2006.

[58] G. Szegö. *Orthogonal Polynomials*. American Mathematical Society, New York, 1939. American Mathematical Society Colloquium Publications, v. 23.

[59] T. Warburton and J. S. Hesthaven. On the constants in *hp*-finite element trace inverse inequalities. *Comput. Methods Appl. Mech. Engrg.*, 192(25):2765–2773, 2003.

[60] J. C.-I. Young and M. J. Crocker. Prediction of transmission loss in mufflers by the finite element method. *J. Acoust. Soc. Am.*, 57(1):144–148, 1975.

# Curriculum Vitae

**Surname**: Parsania
**Firstname**: Asieh
**Date of birth**: August 2nd, 1981, Iran
**Nationality**: Iranian

## Education:

- **Public High School**

September 1995 - July 1999, Iran

- **B.Sc. Applied Mathematics**

Tarbiat Moallem University (TMU), Iran
February 2000 - July 2003

- **M.Sc. Applied Mathematics**

Guilan University, Iran
September 2003 - January 2005
Master Thesis: Surface water Waves involving a Vertical Barrier in the presence of an Ice- Cover.

- **Ph.D. Applied Mathematics**

Since August 2008, Ph.D. student in the Working Group Computational Mathematics of Prof. Dr. Stefan A. Sauter at the University of Zurich